NATURE REVIEWS MICROBIAL OCEANOGRAPHY

nature

THE HOHLE FELS VENUS
Early symbolic art

ORGANIC LEDS
Seeing the light

THE 2008 SICHUAN EARTHQUAKE
Learning the lessons

DAMAGE CONTROL

Small RNAs involved in DNA repair in *Neurospora*

Abstractions

FIRST AUTHOR

When it comes to buying light bulbs, consumer choice is pretty limited. Incandescent bulbs, the most popular option, are widely available and inexpensive, but most of the energy they produce is given off not as light but as heat. Green-minded consumers favour the much more efficient compact fluorescent lamps (CFLs), but CFLs typically cast a colder, less attractive light and, because they contain mercury, are difficult to dispose of. Several groups have thus been working on better alternatives. One option, under study by PhD student Sebastian Reineke and his colleagues at the Institute for Applied Photophysics in Dresden, Germany, is organic light-emitting diodes (OLEDs), thin films of organic molecules that emit light when current passes through them. Until now, their drawbacks have included their comparably low efficiency, but Reineke's team has hit on a solution (see page 234). He tells *Nature* more.

What inspired this work?

Trying to find solutions that save energy has been one of the driving forces of our research. OLEDs had already been shown to have the potential to become one of the next-generation light sources. We are now in global competition to accelerate the commercialization of white OLEDs.

What is the benchmark for energy efficiency?

CFLs provide 60–70 lumens per watt — the ratio of light produced to electricity used — compared with the 15 lumens per watt produced by the average 60-watt incandescent bulb. We have now achieved OLEDs that produce 90 lumens per watt and emit soft area light.

Were there any surprises?

OLEDs emit light as electrical current flows through their organic layers, with the colour of the light depending on the type and number of organic dyes used. But most OLEDs show a shift in colour when the strength of the current passing through them changes — an unwanted feature for dimmable light sources. We were surprised to discover no such colour shift in our OLEDs, no matter how much current we passed through them.

What do you see as future uses for OLEDs?

OLEDs are ultrathin devices that can be scaled to almost any size. You could use one as wallpaper — it would be a thin sheet emitting soft, comfortable light. Or it could be part of a window, where the organic layers — a few hundred nanometres thick, and invisible to the naked eye — are set between layers of glass. During the day it would just look like a window; at night you could turn it on and it would emit light. ■

MAKING THE PAPER

Ami Klin

Point-light animations reveal different focus of those with autism.

Autism is a complex disorder characterized by a lack of social interaction and eye contact, and is typically diagnosed by three years of age. An earlier diagnosis — in the first few months of life — could improve outcome. With this in mind, Ami Klin, a clinical psychologist and director of the autism programme at the Yale School of Medicine, and his colleagues delved into the social development of very young children.

Young animals, from human babies to newly hatched chicks, preferentially focus their attention on the movement of living beings rather than inanimate objects — a trait that enables them to orient to a caregiver, necessary for survival. In 2000, Klin, accompanied by Yale neuroscientist and co-author Warren Jones, went to an animatronic studio in California to create animations that could be used to determine how young children respond to human movement, and how this response might differ in autistic children.

They used a technique that turns videos of human actors playing children's games, such as 'peek-a-boo', into animated dots of light able to convey human motion. Then, by tracking the eye movements of children watching the moving dots, Klin and Jones could measure the children's attention to human movement, and thus social interaction. "The eyes are the window to the soul, but also to socialization," says Klin.

A puzzling observation focused their efforts. A 15-month-old girl whose brother had autism was shown a screen divided in two: on one side the light displays were upright; on the other they were inverted, so no longer representative of human movement, and played backwards. The girl showed no preference for upright or inverted images — with one exception. During the 'pat-a-cake' video, one of the nine animations presented to her, she focused almost entirely on the upright video. "At first, we were confused," says Klin. However, further investigation established that this was the only animation in which the sound — in this case, clapping — was clearly synchronized to the light movement.

The duo suspected that autistic children might be more attentive to a physical stimulus (sound synchronized to motion) than to a social one (human movement). To test the idea, they showed the animations to groups of 2-year-old children with and without autism, and found that only those with autism had the same response as the 15-month-old girl (see page 257). "Then we sat back and thought we should be adventurous in order to learn the profound lesson this little girl was teaching us," says Klin.

They and their colleagues at Yale spent the next two years coming up with a method to quantify how much audiovisual synchrony there was in different animations — and compared that measurement with children's visual behaviour. In the end they were able to predict, with 90% accuracy, the children's visual preference on the basis of even incidental bits of audiovisual synchrony present in the animations.

Klin says the little girl helped them to understand that autistic children develop in a world where preferential attention is given to physical, rather than social, attributes. On the basis of this realization, Klin and Jones are now looking at ways to pinpoint when this divergence from typical social development begins. "We want to come up with a behaviour assay that will predictably diagnose vulnerabilities for autism in the first year, if not months, of life," Klin says. ■

FROM THE BLOGOSPHERE

Women scientists in India get some inspiration to go with their aspirations. Writing on the Indigenus blog, Nature India editor Subhra Priyadarshini highlights a new book, *Lilavati's Daughters: The Women Scientists of India* (http://tinyurl.com/pqgknz). The book, named for the legendary daughter of a twelfth-century mathematician, presents 98 biographical essays and Priyadarshini recommends it for those in search of role models.

The book "has every emotion one ever attributes to women scientists — patience, angst, perseverance, fears, euphoria and above all incessant struggle in the face of a thousand odds," she writes. The book comes on the heels of announcements of government programmes aimed at easing the burdens of family-work balance on young women to help stem the high numbers dropping out of science. Whether such programmes will be implemented properly or embraced by women scientists is a topic that has been raised in several discussions at the Nature India forum.

The Indigenus post includes a link to the Indian Academy of Sciences, where the book can be read online (http://tinyurl.com/qtc5h7). ■

Visit Nautilus for regular news relevant to *Nature* authors ▸ http://blogs.nature.com/nautilus and see Peer-to-Peer for news for peer reviewers and about peer review ▸ http://blogs.nature.com/peer-to-peer.

nature

www.nature.com/nature

Vol 459 | Issue no. 7244 | 14 May 2009

Politics proves its worth

The European Parliament has reaffirmed its legislative value by reversing the potentially disruptive restrictions in the draft directive for protecting laboratory animals.

Raise a glass to the elected members of the European Parliament (MEPs). Without their intervention last week, the European Union (EU) directive on the protection of laboratory animals would have continued its tortured path through legislative procedures in a form that was thoroughly toxic to biomedical research.

The European Commission, the EU's executive arm, began working on the directive back in 2002.The draft that finally emerged last November was singularly uninformed. It should have balanced the undisputed duty to protect animals with the needs of biomedical researchers to understand disease and develop innovative therapies. Instead, it proposed restrictions that would have blocked whole areas of fundamental research while having no positive influence on animal welfare. In particular, it would have restricted research on non-human primates to "life-threatening or debilitating diseases" (which it did not define) without thought for the basic research required to understand such diseases biologically.

The draft would also have forbidden the reuse of animals in any procedure that could cause more than a "mild" (again, undefined) level of suffering. This would have ruled out the use of surgically implanted telemetric devices, which continuously monitor physiological aspects such as blood pressure or heart rate, save animals the stress of frequent handling, and allow for the testing of different compounds on a single animal. As surgery could hardly be classified as "mild", an animal would have to be killed after just a single test.

Justifiably alarmed, researchers (and *Nature*, see **456,** 281–282; 2008) added their voices to the powerful lobby of the European drug industry — and MEPs responded. In last week's vote, the European Parliament reversed most of the problematic clauses.

So why did things go so wrong at the commission? The legislation was handled in its environment directorate, which initially consulted with all the stake-holders, but then shut itself off from all influences except the powerful animal-welfare lobby. It even failed to consult on the text with the research directorate. Then, when the text was at last opened to the entire commission for comment last summer, there was little time to make substantial changes.

In parliament, by contrast, the procedure was transparent and professional. The draft directive was examined by three committees — agriculture, research and environment— which considered the interests of animal welfare and researchers with appropriate balance.

The process is far from over, however. According to Europe's elaborate co-decision process, not only the parliament, which is directly elected every five years, but also the European Council of Ministers, comprising representatives of each of the EU's 27 member states, must agree on the final text in two readings. The council will start work on the amended text during the Swedish presidency, which begins in July. The commission will then redraft the directive, taking into account the wishes of parliament and council before the second reading. Changes can be introduced at any stage. But in the final directive, which is likely to be approved during 2011, the interests of research will not be as neglected as they were at the outset.

The European Parliament is the only one of the three EU bodies that is elected and therefore directly answerable to EU citizens. This example shows how important it is to have research-savvy MEPs. The next election takes place early next month. Scientists in the EU would be well advised to consider their local candidates' attitude towards science and to cast their vote accordingly. Meanwhile, researchers and their organizations should keep their eye on the passage of the directive, and keep their campaign weaponry close at hand. ■

Bracing for the unknown

Last year's earthquake in China is a salutary reminder about preparing for risk in the face of uncertainty.

Despite a century of research into earthquakes, Earth scientists are still only beginning to understand how individual faults behave. Although many dangerous faults have been identified, which has helped countries to strengthen their infrastructure, a significant number of deadly earthquakes occur on faults that are either unknown or were not thought to be particularly dangerous. That knowledge gap was highlighted last year, when a group of faults not particularly high on China's list of hazards linked together in an unexpected manner to spawn one of the most deadly quakes in recorded history, claiming at least 70,000 lives in Sichuan province (see page 153).

Earthquakes clearly pose the problem of how to prepare for risk in the face of uncertainty. The answer is complex, but can be boiled down to a few fundamental principles that scientists and government leaders should take to heart. Develop a clear message about what is known and — just as importantly — what is unknown. Be forthcoming about mistakes. And use a broad set of tools to prepare for hazards — a strategy that will make communities more resilient to different kinds of threat.

Scientists must rigorously assess the limits of their knowledge and communicate them to officials and the public. Earthquake researchers in some regions are getting better at this. California, for example, is one of the best-studied regions in terms of seismic risk. Two decades ago, seismologists there began issuing semi-regular reports on the major threats. Early on, they adopted a relatively rigid approach based on the understanding that segments of the San Andreas fault tended to behave in certain set ways, with characteristically sized earthquakes. But over time, the data — and the reports based on them

— have grown less definite. The most recent assessment, released last year, acknowledges the complexity and uncertainty of fault behaviour more than past reports.

As for public officials, they must admit their mistakes and seek to learn from them — a lesson powerfully demonstrated during America's bungled response to Hurricane Katrina in 2005. In Sichuan, a large number of schools collapsed in the quake zone and too few answers have been offered by political leaders there about what happened. Amnesty International reported this month, for example, that the Chinese authorities have detained parents who have demanded information about the collapsed schools that killed their children.

The Chinese government must be forthcoming about what happened if it and other countries are to learn from this incident. Engineers who toured the site noted that some types of school building along one of the involved faults did not collapse whereas many others did. Data about school construction would clearly help to save lives in future disasters: the survival of some schools shows that structures can be designed to withstand severe quakes even in regions with limited resources.

Scientists, government officials and the public must strive to make societies more resilient to earthquakes and other natural hazards. Social-science research shows that citizens are generally poor judges of the hazards they face: they think they are safe until disaster strikes. The obvious but difficult truth is that societies must prepare for disasters before they occur. That means raising public awareness of the need to do so, something that Japan accomplishes with its annual earthquake drill each September. California last year successfully staged its first such drill and is planning to repeat it in October.

"Citizens are generally poor judges of the hazards they face."

With public support, government officials can guard against earthquake losses by taking a multipronged approach. Buildings codes and land-use regulations — when rigorously enforced — can make structures safer. And societies can improve their ability to respond to quakes by strengthening their emergency systems as well as their capacity for reconstruction. Such preparations will also help nations to weather terrorist attacks, climate change and many of the other threats present on this dangerous planet. ■

A measure of marine life

The extraordinary emerging images of ocean microbiology need the fourth dimension of time.

The globe floating in the void might almost be the first, haunting glimpse of an alien world, blue-green and dense with life. But this is the view through a microscope, not a telescope, and the globe is a crucial inhabitant of this planet, not a token of another. Just a micrometre in circumference, *Prochlorococcus* makes up for in number what it lacks in size. This tiny bacterium is the most common photosynthetic organism on Earth, providing a substantial fraction of the planet's carbon fixation.

Until just over two decades ago, moreover, *Prochlorococcus* was unknown. Its ubiquitous presence in non-coastal, non-polar waters is one of many recent discoveries by which ocean microbiology has re-emphasized the primacy of microbes in Earth's biosphere. That primacy, analysed in this week's Insight starting on page 179, holds everywhere, notably in soils. But in soils, every pore and grain provides its own microenvironment. The seas, transparent and well mixed, are where this microbe-centric view of life is most clearly visible.

The new discoveries have revealed a previously unimagined profusion of microbial life, with perhaps 1,000 times more organisms per unit volume than scientists thought in the 1980s. Not only are there new players, such as *Prochlorococcus*, but whole new classes of player, such as the Archaea now known to exist far more widely than suggested by their early reputation as niche extremophiles. There are metabolisms that were previously unknown, such as those of bacteria using sunlight and proteorhodopsin to power their lives. And, perhaps most exciting of all, there is an extraordinary amount of genetic diversity and gene transfer — the latter often mediated by unexpectedly abundant viruses.

These findings are largely based on improved technologies, such as satellite imaging that can read out chlorophyll abundances, flow cytometry that can distinguish the tiniest cells, and gene sequencing that can make sense of raw genomes from the water. Between them they have provided a picture of life in the oceans covering every scale from the pigment molecule to the planet as a whole. Yet for all their data-gathering power, these technologies are still largely blind to the temporal dimension — a problem that urgently needs addressing.

The oceans, after all, are patterned in time as well as space. The Hawaii Ocean Time-Series, running now for two decades, has seen intriguing signs of long-term oscillations between nitrogen and phosphate-limited microbial assemblages. But such thorough, regularly assessed measurements of the physical, chemical and biological environment in the oceans are almost nonexistent.

To reproduce such time series in dozens of ecologically and oceanographically distinct provinces around the world would be costly, and hard to justify on the basis of traditional hypothesis-driven science. But that is not the correct yardstick for this work. At a time when humankind's carbon emissions are producing rapid changes in Earth's climate, recording those changes as they reverberate through the seas is a necessity if they are to be understood, and their future course predicted.

"New discoveries have revealed a previously unimagined profusion of microbial life."

More generally, there is a growing number of areas in which scientists' ability to gather information currently exceeds their ability to understand it. Although it may go against the grain, it is worth considering that gathering those data regardless of comprehension is worth some effort, even if there are opportunity costs to current science. The scientists of the future, with greater knowledge, craft and insight, will be grateful for them.

This argument applies particularly strongly to attempts to monitor Earth and its oceans as a whole. Unlike the photosynthetic galaxies strewn across the seas, this planet is the only available example of its type for humanity to understand. The light-rich, life-rich seas are a key to that understanding. ■

RESEARCH HIGHLIGHTS

The long bask

Curr. Biol. doi:10.1016/j.cub.2009.04.019 (2009)
The question of where basking sharks — the world's second largest fish — in the western Atlantic go in winter has been answered.

Gregory Skomal of the Massachusetts Division of Marine Fisheries in Oak Bluffs and his colleagues tagged 25 basking sharks (*Cetorhinus maximus*) with temperature, depth and light-level recorders that popped off after a given interval. Reconstructing six of the creatures' travels, the researchers found that the sharks covered distances of about 9,000 kilometres and dived to depths of up to 1,000 metres, heading to deep tropical waters in the winter.

The sharks were formerly thought to be restricted to temperate waters, and the researchers are not sure why they travel so far. Perhaps, they speculate, their young are born deep in the tropics.

For a longer story on this research, see http://tinyurl.com/pwusgt.

L. PITKIN/NHPA

ECOLOGY

Bouillabaisse

Glob. Change Biol. doi:10.1111/j.1365-2486.2009.01875.x (2009)
A study of larvae of fishes off southern California has shown for the first time how climate change can affect the distribution and abundance of species.

Chih-Hao Hsieh, now at the National Taiwan University in Taipei, and his colleagues studied 34 species. When comparing data from a cooler period of 1951–1976 with those from a warmer time of 1977–1998, the team found a significant shift in the vertical or lateral distribution of 16 species, and that eight species had shifted their larger geographical distribution. The plankton-eating fishes typically sought cooler waters.

Surprisingly, the group found an overall increase in abundance, and that offshore fishes moved closer to shore. Thus climate change can drive species into new habitats, which could have unexpected ecological consequences.

ASTRONOMY

Strange star

Astrophys. J. **697,** L63–L67 (2009)
Astronomers have spotted a star with an unique mix of chemical elements in the Milky Way's halo, suggesting that stars in the Galaxy's outer reaches are more varied than previously believed.

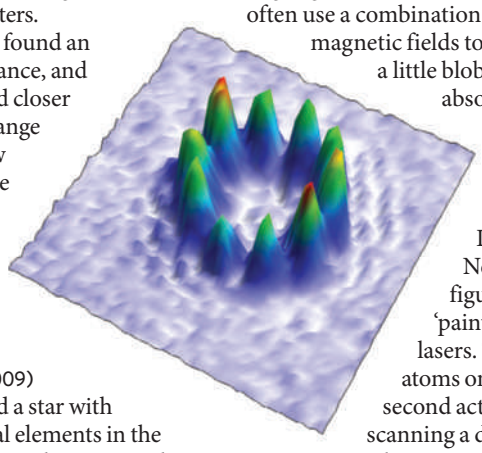
David Lai of the University of California, Santa Cruz, and his colleagues studied 27 stars in the Galactic outer halo, some 50,000 light years from Earth and beyond. Spectral analysis of one star's light showed that it contains high amounts of calcium relative to other elements such as iron and magnesium.

The authors say that the star may have been accreted into the outer halo from another nearby star system, suggesting that the Galaxy's history is more dynamic than thought.

QUANTUM PHYSICS

Atomic painting

N. J. Phys. **11,** 043030 (2009)
Bose–Einstein condensates (BECs) are clouds of ultracold atoms that behave as a single, giant quantum object. Physicists often use a combination of laser light and magnetic fields to trap and then cool a little blob of atoms to near absolute zero.



Malcolm Boshier and his colleagues at Los Alamos National Laboratory in New Mexico have figured out how to 'paint' a BEC using two lasers. The first traps the atoms on a flat canvas; the second acts as a paintbrush, scanning a desired shape and cooling it until a BEC forms. The group can make a BEC of rubidium atoms in any shape (example pictured left), for use in fundamental studies or quantum information processing.

M. BOSHIER

GEOSCIENCES

The forever landscape

GSA Bulletin **121,** 688–697 (2009)
The rough surface of Israel's Negev Desert has the slowest rates of erosion ever measured, according to Ari Matmon of the Hebrew University of Jerusalem and his colleagues. His team calculated the speed of erosion there by measuring the concentration of the radioactive isotope beryllium-10 in chert clasts — little stones — collected from sites in the Negev. This isotope is formed when cosmic rays hit rocks and soils, so its concentration can indicate how long an object has been exposed to the sky.

This technique, along with others, suggests that the bits of chert covering parts of the Negev, Sinai, Sahara and Arabian deserts have been sitting there for upwards of 2 million years, making this landform the longest-lived one on Earth according to current measurements.

IMAGING

Seeing beyond skin deep

Science **324,** 804–807 (2009)
A team led by Roger Tsien of the University of California, San Diego, reports that it has engineered the first protein that emits infrared light and can be used to image intact animals.

Existing fluorescent proteins are excited by shorter wavelengths, which don't penetrate far into animals' bodies. The new proteins were made from a light-detecting pigment called a phytochrome from the bacteria *Deinococcus radiodurans*. The phytochrome

naturally incorporates a green pigment, biliverdin, that is abundant in animal tissues. Tsien's team modified the phytochrome so that it rigidifies biliverdin, which then absorbs far-red light and emits infrared light.

The researchers showed that the modified phytochrome can be used to image an animal's inner tissues, such as the liver, and say that it could be useful in fields such as cancer and stem-cell research.

MATERIALS

Everlasting memory

Nano Lett. doi: 10.1021/nl803800c (2009)

The data packed as magnetic regions on hard disks will fade in just a few decades, as atoms vibrate and reorient themselves.

But an iron nanoparticle sheathed inside a carbon nanotube could form a protected data element, whose position would remain stable at room temperature for more than a billion years, report Alex Zettl of the University of California, Berkeley, and his team.

By applying an electric pulse, the researchers controllably shift the nanoparticle back and forth. Its position — corresponding to a '0' or a '1' — can be easily read by measuring electrical resistance across the nanotube.

A device made of bundles of individually positionable nanotubes could form an ultra-high-density data store, readable for any practical time scale, the researchers think.

POLYMER CHEMISTRY

Doughnut machine

Angew. Chem. Int. Edn doi:10.1002/anie.200900533 (2009)

In solution, block copolymers — different types of synthetic polymer linked together — spontaneously cluster into a dazzling variety of shapes, including spheres, cylinders, discs and helices. Lately, even ring doughnuts (toroids) have been observed — but never alone, and always of varying size.

Taihyun Chang and his colleagues at Pohang University of Science and Technology in Korea have now hit on a recipe of copolymer and solvent that for the first time produces pure, almost uniform toroids — all about 70 nanometres in diameter and with a ring about 30 nanometres thick in cross-section. They are stable in solution for several months.

It is not clear how these doughnuts form; potential applications include use as templates for nanometre-scale patterning. For example, the researchers use them as a template to grow rings of gold nanoparticles around the doughnuts' edges.

MICROBIOLOGY

On the surface

PLoS Pathog. **5,** e1000407 (2009)

The bacterium associated with stomach ulcers creates a habitable environment by clinging to human cells and interfering with their polarity.

Helicobacter pylori avoids the stomach's lethal acidity by colonizing a thin layer of mucus that coats stomach epithelial cells. These cells are polarized — that is, the outside surface facing the stomach and the inside surface, which backs onto the underlying tissue, have different properties.

Manuel Amieva and his colleagues at Stanford University in California found that the bacterium can thrive when attached to these cells in culture, even when the culture medium lacks nutrients normally required for survival. However, *H. pylori* mutants that lacked a protein called CagA were not able to colonize the outside surface of these cells. CagA is known to alter the polarity of epithelial cells, presumably making the outside surface of the cells more like the inside surface, and thus making them colonizable.

M. VENCES

CONSERVATION

Amphibian additions

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0810821106 (2009)

Madagascar is a biodiversity hotspot but, according to David Vieites of the Spanish National Research Council (CSIC) in Madrid and his colleagues, it may be even hotter than we think.

They sequenced the DNA of 2,850 amphibian specimens collected from more than 170 locations on the island. Analysis of these sequences suggests that at least another 129 amphibians remain to be described on Madagascar, including the frog pictured above.

At a maximum, the authors say, there may be 221 species missing from current records. This would represent an increase of almost 100% on the 244 described so far and an increase of 250% since 1991.

JOURNAL CLUB

Lee Turnpenny
University of Southampton, UK

A stem-cell researcher considers an accusation of dullness.

How might hard-working scientists react to an accusation that 'modern scientists' are 'dull', as is provocatively postulated in a March editorial of the non-peer-reviewed journal *Medical Hypotheses* (B. Charlton *Med. Hypotheses* **72,** 237–243; 2009). With offence? Humour? Ambivalence? Or, perhaps, in response to a jeremiad bemoaning our apparent insufficient intelligence and creativity, we might retort, "So what? Tell us something we don't know."

Because, it seems to me, most working scientists have either long since accepted that they are not of the 'revolutionary' type exemplified by greats such as Isaac Newton, Charles Darwin and Albert Einstein, or never strived to be. Gaining and retaining employment in academia is hard enough. Yes, we are of the persevering and conscientious 'normal' type — if we weren't, nothing would get done.

We know there is too much bureaucracy. And yes, there is a lot of repetitive, boring, tiresome, problematic work to be done that is unlikely to shift any paradigms (yet), but important nonetheless. Whether or not somehow creating more windows of opportunity for would-be geniuses possessed of the requisite levels of selfishness and creativity would lead to significant changes in direction is debatable. But the drudge is always necessary in a multidisciplinary collaborative enterprise.

It's not that scientists are dull per se. Rather, instead of being the 'clever crazy' type that might belong in an institution, we labour in an institutionalized occupation that demands we play by certain rules. We know we're not going to change the world, but we like to think we can contribute to the sum of knowledge. Providing we can first convince our peers. If it was easy, everybody would do it. One might add, complaining that modern science can be dull, although valid, isn't exactly a 'revolutionary' idea. Tell us something original, eh?

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

Vaccine decisions loom for new flu strain

World Health Organization considers live attenuated vaccines for swine-associated H1N1 outbreak.

Faced with the prospect of an influenza pandemic, the World Health Organization (WHO) is weighing up its options for advising manufacturers and governments on developing vaccines. With current manufacturing capabilities, there will be enough vaccine for only a fraction of the world's population, and not before six months from now. And most of that will go to rich countries.

Experts are meeting at the WHO on 14 May to discuss options for proceeding with a vaccine for the currently circulating swine-associated H1N1 strain. One controversial idea being floated is to use a live attenuated vaccine, which could boost the number of doses available from existing plants by 50- to 100-fold.

"Everybody is anxious to have enough seasonal vaccine."

Manufacturers are lukewarm to the idea. The ordinary seasonal flu vaccine uses inactivated virus, and serious regulatory barriers exist to introducing a live-virus vaccine. Demonstrating efficacy and getting regulatory approval in time would pose "quite significant difficulties", says George Kemble, vice-president of vaccine research and development at MedImmune in Gaithersburg, Maryland, which makes live flu-virus vaccine.

But some experts say the live-virus idea should be entertained. Two factors largely determine whether a vaccine can protect large numbers of people during a pandemic. One is the delay before substantial quantities of vaccine become available — usually around six months, due to the time required to grow the virus in hens' eggs. (Cell culture and other technologies could be faster, but they are not yet ready for prime time.) The second limiting factor is production capacity, currently at around 700 million to 900 million doses of seasonal flu vaccine annually.

Although still limited, production capacity is much better than five years ago — when it was around 300 million — mainly because of measures taken by governments to prepare for a pandemic threat. The seasonal vaccine contains antigens against three circulating flu strains. Switching to producing a single vaccine against just the new H1N1 virus could, in principle, mean that existing plants could make three times as many doses.

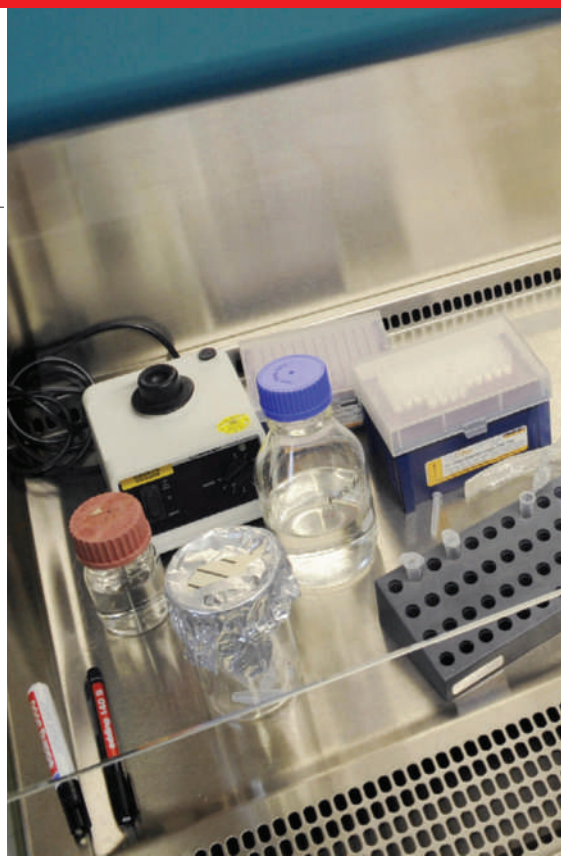
But even if global facilities switched entirely to producing an inactivated H1N1 vaccine, only about 1 billion doses at most are expected to be available by the end of the year, around the time of the Northern Hemisphere flu season. Moreover, because the population has little to no pre-existing immunity, the vaccine will probably need to be given in two doses — reducing the actual number of vaccines to 500 million.

Switching even part of the production to a live-virus vaccine would effectively increase production capacity. The virus in such a vaccine is capable of reproducing in humans, so much lower doses can be given. Live-virus vaccines also don't require adjuvants to bolster their effectiveness, can be administered nasally — avoiding the need for syringes — and are thought to provoke a broader and stronger immune response than inactivated vaccine. And whereas one egg yields one dose of inactive vaccine, for a live vaccine it yields somewhere from 50 to 100 doses.

Whatever routes are taken, production capacity will also depend on how well the new H1N1 virus can be grown and cultivated. The news here seems to be good. "We've done 7 cycles, 42 hours each, and it's going very well," Doris Bucher, an immunologist at New York Medical College, told *Nature*. The US Centers for Disease Control and Prevention in Atlanta, Georgia, asked her to help grow the first reference strains to be sent to manufacturers. The immune response produced by the resultant seed vaccines will need to be tested in clinical trials; if it were, for example, to require three times as much antigen as seasonal flu to prompt an adequate immune response, that would cut theoretical production capacity to a third.

To grow live attenuated vaccines, scientists would reassort the new flu strain with a 25 °C cold-adapted strain, which will multiply in the nose but not grow in the higher temperatures of the lower respiratory tract.

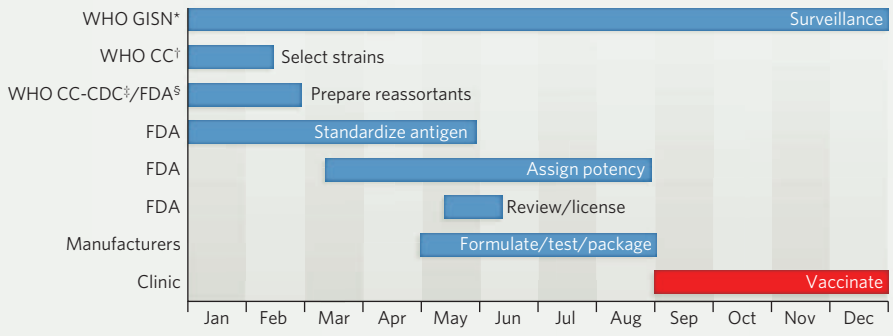
At present, only two groups have the technology to produce live attenuated flu vaccines:



T. LOHNES/AFP/GETTY IMAGES

THE PRODUCTION CYCLE

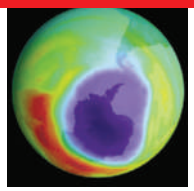
Rushing a swine flu vaccine is difficult; this timeline, using the United States as an example, illustrates how vaccine production takes at least six months from selecting a strain to producing the vaccine.



*World Health Organization Global Influenza Surveillance Network †WHO Collaborating Centres ‡US Centers for Disease Control and Prevention §US Food and Drug Administration

Source: CDC

Culturing vaccines in chicken eggs is the main time delay in production.



NASA

OZONE ARGUMENT RESOLVED
Chlorine radicals on best behaviour.
www.nature.com/news

MedImmune, and Nobilon, a subsidiary of Schering-Plough, which has licensed technology developed at the Institute of Experimental Medicine in St Petersburg, Russia. MedImmune's FluMist is approved for use in the United States for those aged 2-49 years old; older people have been exposed to past pandemic viruses, and their immune systems therefore kill the live vaccine for current circulating strains. The WHO has obtained a licence from Nobilon to allow manufacturers in developing countries to use the Russian technology.

Inactivated vaccine makers seem sceptical about using live attenuated vaccines more widely, even in a pandemic situation. Changing over to a live vaccine would mean introducing new production methods and possibly having to license outside technology, such as that from MedImmune. There are also safety and liability issues, because it would be difficult to organize clinical trials of an untested vaccine quickly enough.

"Given the potential level of global needs, particularly if the threat worsens, all serious vaccine candidates and approaches should be actively considered and vaccine candidates made ready, including live attenuated vaccine," says Jesse Goodman, acting chief scientist of the US Food and Drug Administration. He notes that live attenuated vaccine is approved in the United States for children and young adults, who "may be at particular risk of infection" in the current H1N1 outbreak. Safety, however, is paramount, he adds: "It is also important to keep in mind, even in the face of a pandemic threat, the importance of doing all that is appropriate and possible to assure high vaccine quality and safety, particularly if and as new facilities, processes and products may be considered."

Indeed, memories are still vivid of the 1976 flu-vaccine fiasco. That year, a new swine flu emerged at an army barracks in New Jersey, killing one person but failing to spread further. A mass vaccination campaign ordered by president Gerald Ford caused neurological side effects in some people, and killed 25.

David Fedson, a pandemic-vaccine expert and retired former medical director of French vaccine development company Aventis-Pasteur, now known as Sanofi Pasteur, says that companies are far more comfortable working with inactivated viruses. But he argues that the live-virus approach is the way to go now with H1N1. "Clinical trials of live vaccine have shown it to be safe in people up to 50 years of age," he says. "What we're talking about here is a monovalent, and hence simpler, vaccine with a new H1N1 virus replacing an older H1N1 virus."

H1N1 update

As confirmed by the World Health Organization, as of 12 May:
Cases: 5,251, in 30 countries
Deaths: 61, in 4 countries

Whatever the WHO decides to do, it will have to balance a switch from a trivalent seasonal flu to a monovalent pandemic vaccine, with the need to make and distribute enough seasonal vaccine in time for the next flu seasons in both hemispheres (see graphic). The WHO is currently surveying all vaccine manufacturers — the bulk of whom are based in Europe — to ascertain where they are in that process, and how ready they would be to switch to a monovalent vaccine.

That they will at some point is a foregone conclusion. Initial information suggests that companies are well along in producing the seasonal H1N1 and H3N2 strains for the Northern Hemisphere, but are having difficulty growing the third influenza B strain. One option may be to go with just the two antigens under way, and drop the B from next year's vaccine. Northern Hemisphere production would then be freed up faster to work on a swine-flu vaccine.

Another question is how to dovetail that with the needs of the Southern Hemisphere, where vaccine production typically starts around November and continues through March. The amounts of vaccine ordered by Southern Hemisphere countries is much lower than that by the north, meaning that northern manufacturers might then have extra time to work on a swine-flu vaccine. A decision on whether to curtail Southern Hemisphere production does not need to be made for weeks or even months, says Marie-Paule Kieny, head of the WHO's Initiative for Vaccine Research.

The WHO is not yet asking its advisory panel to consider calling a halt to seasonal flu-vaccine production. What they are likely to recommend this week is whether or not companies should go to commercial-scale production of a monovalent pandemic vaccine at the earliest opportunity. "Everybody is anxious to have enough seasonal vaccine," says Kieny. One scenario, she says, is that full-scale manufacturing of swine-flu vaccine could start by July at the earliest.

Margaret Chan, director-general of the WHO, will meet on 19 May with the heads of flu-vaccine companies to discuss ways forward, and how developing countries could access the vaccine. Kieny says that experience with mass vaccination campaigns in developing countries, for example for meningitis, should make it easy to quickly deploy a vaccine.

CDC

In the meantime, the new H1N1 strain remains susceptible to the antiviral drugs oseltamivir (Tamiflu) and zanamivir (Relenza). Cheaper and more widely available antibiotics and anti-inflammatories, such as statins, could also have a role in limiting mortality in a severe pandemic. "We should be giving attention to looking at new ways of treatment," says Kieny. "We must not give people the impression that if they don't get a vaccine or Tamiflu that they will die." ■

Declan Butler

For *Nature*'s ongoing coverage of the H1N1 outbreak, including a Q&A with virus grower Doris Bucher, see www.nature.com/swineflu.

Stem-cell therapy faces more scrutiny in China

But regulations remain unclear for companies that supply treatments.

CNIMAGING/NEWSCOM

BEIJING

The Chinese Ministry of Health has implemented regulations on the clinical application of cutting-edge therapies such as stem-cell injections.

Stem-cell scientists in China contacted by *Nature* hope that the rules may help to curtail a growing trade in unproven treatments that attract patients from around the world, risking their health and potentially damaging the reputation of stem-cell research.

The new regulations, which came into effect on 1 May, designate all forms of stem-cell therapy as 'category 3' medical technologies — those deemed "ethically problematic", "high risk" or "still in need of clinical verification". The ministry will take direct responsibility for regulating all category-3 procedures, which include gene therapy, surgical treatment of mental disorders or drug addiction, and sex changes.

Institutions wishing to offer stem-cell therapies must first demonstrate safety and efficacy in clinical trials; the treatment will then be assessed by a ministry-approved regulator. Institutions failing that process must wait 12 months before reapplying. Although the penalties for not adhering to these rules have not been made explicit, institutions that transgress are likely to face fines or have their permit to practice medicine revoked, says Renzong Qiu, a bioethicist based at the Peking Union Medical College in Beijing.

"These regulations will make people understand that the Ministry of Health and many scientists in China are concerned about these unverified procedures," says Ching-Li Hu, a paediatrician and senior adviser to Shanghai Jiaotong University's medical school, and a member of the International Bioethics Committee of the United Nations Educational, Scientific and Cultural Organization.

Hu and Qiu are members of an expert panel that will deliver recommendations to the ministry later this year on how to implement the regulations effectively.

Murky area

China already has experience in regulating cutting-edge technologies by assessing clinical trials and conducting ethical reviews. It was the first country to give governmental approval for a gene-therapy treatment, one produced by SiBiono GeneTech in Shenzhen that targets head and neck cancers.

But stem-cell therapy is a murkier area. Some researchers worry that medical institutions will be able to circumvent the regulations by calling their therapies research, even though they are charging patients and not carrying out the rigorous monitoring required by clinical-trial protocols. If those institutions have sought official approval, it comes from local governments or institutional review boards, which do not have the expertise to properly assess the treatment, says Hu.

From interviews with scientists and physicians, Qiu estimates that there are 100–150 clinics claiming to offer stem-cell therapies in China. But it is not yet clear whether companies supplying the stem cells will be also be

Exome sequencing takes centre stage in cancer profiling

COLD SPRING HARBOR

To help battle their way through the stream of data coming in from human gene sequencing, major cancer-genome screening projects such as the International Cancer Genome Consortium (ICGC) seem to be choosing to simplify matters.

The ICGC aims eventually to sequence the full genomes of 25,000 tumour samples as well as those of the people from whom the tumours were taken, which would give 50,000 distinct genomes.

But in the near term, the project is doing targeted sequencing of just the 1% of the genome known to code for proteins — the 'exons' within genes.

Sequencing of the 'exome'—all the exons in the genome—involves chopping the genome into millions of pieces and capturing and sequencing only selected DNA from exon regions. It differs from transcriptome sequencing by focusing on DNA rather than the expressed RNA in a given cell, and it promises to be vastly cheaper than whole-genome sequencing. It will be a significant focus of the ICGC, which comprises ten projects from nine member countries, says Tom Hudson of the Ontario Institute for Cancer Research in Toronto and a member of the ICGC secretariat.

Last week, at the 'Biology of Genomes' meeting at Cold Spring Harbor Laboratory in New York state, some cancer researchers questioned whether exome sequencing is the most efficient way forward. They say it could represent a piecemeal half-step, and not provide a full picture of the mutations that lead to cancer.

At the conference, Michael Stratton of the Wellcome Trust Sanger Institute in Cambridge, UK, presented early results from a study of 24 breast-cancer samples that analysed more than 2,000 chromosomal rearrangements, including regions in which vast tracts of DNA were duplicated, swapped between chromosomes, inverted or otherwise adulterated.

With so many potentially deleterious rearrangements occurring in any given cancer cell, it becomes difficult to distinguish what Stratton calls the "driver" mutations, which spur and maintain cancer development, from "passenger" mutations that are just along for the ride.

Drivers might be in the coding regions of the genome, but some will presumably be in regulatory elements and other non-coding

Experts estimate that stem-cell treatments are offered by more than one hundred clinics across China.

subject to the regulations.

Shenzhen-based Beike Biotechnology is China's most prominent stem-cell therapy company, providing adult stem cells and umbilical-cord stem cells to a network of 27 clinics worldwide. The company also acts as a first point of contact for patients. Luca Ricci, the Beike representative at Zhejiang Xiaoshan Hospital in Hangzhou, told *Nature* that his job was to "work in the hospital as an interpreter, taking care of the patient before and after they arrive." Beike's medical officer, Kara Zhang, says that she visits patients to provide medical consultations.

The company claims that more than 4,000 patients have been treated for disorders including autism, cerebral palsy, multiple sclerosis and spinal-cord injury. Over the past year, several media reports have claimed that the company's stem-cell treatments have restored sight to blind children.

But the treatments have not been subject to controlled clinical trials to assess whether they are effective and safe — and they don't come cheap. Earlier this year, Beike quoted a price of US$26,300 for an initial course of six stem-cell injections to treat a patient with spinal muscular atrophy, with additional injections costing $3,500 each.

"Having the company that provides the cells interacting directly with patients at an independent hospital or institution should be prohibited," argues David Magnus, director of the Stanford Center for Biomedical Ethics in California. In his opinion, the situation seems to be "the equivalent of a drug rep selling an unproven product directly to the patients at the hospital."

Beike did not answer *Nature's* questions about the scientific evidence supporting its stem-cell treatments; their success rates; their reaction to the ministry's regulations; whether they had published any results from their procedures in a peer-reviewed journal; or whether they had conducted any clinical trials. But the company has certainly considered clinical trials. In early 2008, Beike and the Minneapolis Heart Institute Foundation in Minnesota discussed jointly pursuing clinical trials on using stem cells to mitigate certain heart disorders.

The foundation offered to help Beike set up a clinical-trial protocol that would include creating a registry of patient outcomes. Joseph Cosico, the foundation's vice-president for research operations, says that Beike declined the offer "because of their inability to fund the venture". Beike says that it decided to work with another group, partly for cost reasons, but would not provide any details of that collaboration.

"I can understand why they wouldn't want to do a trial," says cell biologist Duanqing Pei, director-general of the Guangzhou Institute of Biomedicine and Health. "They might spend millions of dollars to prove that the treatment isn't effective." ■

David Cyranoski

sequences — meaning that whole-genome sequences will ultimately be necessary, he says.

Elaine Mardis, of Washington University in St Louis, Missouri, offered a glimpse of what else could get missed by focusing on the exome with current technologies. Building on her recent whole-genome sequences of a patient with acute myeloid leukaemia (*Nature* 456, 66–72; 2008), she presented data on a second patient–tumour pair and preliminary data on a third. With hundreds of potential mutations churned up everywhere in the genome, her group focused on validating three different 'tiers' of single-nucleotide mutations, many of which lie in coding regions.

Asked why non-coding elements hadn't got more attention, she replied that her group was looking at these regions but that they would need more work to sort out, hopefully with the help of expression data and comparison with other patients. "Right now, it's not worth it," she said.

Nevertheless, Mardis remains a big fan of sequencing whole genomes. She says that the exome approach, which uses new techniques to capture the targeted DNA for sequencing, can miss as much as 20% of the coding regions.

"If the amount of data is scary, why not sequence the whole genome and then just focus on the genes?" she asks. "You could posit that is ultimately a cheaper approach than trying to get 100% of the genes, only coming up with 80%, and then going to some extraordinary measures to get the remainder that you missed."

Francis Collins, former head of the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, agrees. "None of the methods are perfect," he says. But he predicts that in the near future, "exome sequencing is where most of the action is going to be".

And many cancer researchers see exome sequencing as a reasonable stop-gap solution until sequencing whole genomes becomes cheap enough. Lynda Chin of the Dana-Farber Cancer Institute in Boston, Massachusetts, says that exomes are a faster way in to identifying driver genes, and help accelerate better screening methods and treatments.

Chin has headed up some of the projects for the Cancer Genome Atlas (TCGA), a potentially billion-dollar-plus programme announced in 2005, and directed by the NHGRI and the National Cancer Institute. TCGA is now moving out of its pilot phase, in which it sequenced and characterized hundreds of tumours from three different types of cancer found in lungs, ovaries and brain, towards characterizing 20–25 cancer types.

In conjunction with some whole-genome sequencing, says Chin, exome sequencing will be part of the new portfolio. "We have to push the envelope," she says, "now". ■

Brendan Maher

Deep concerns

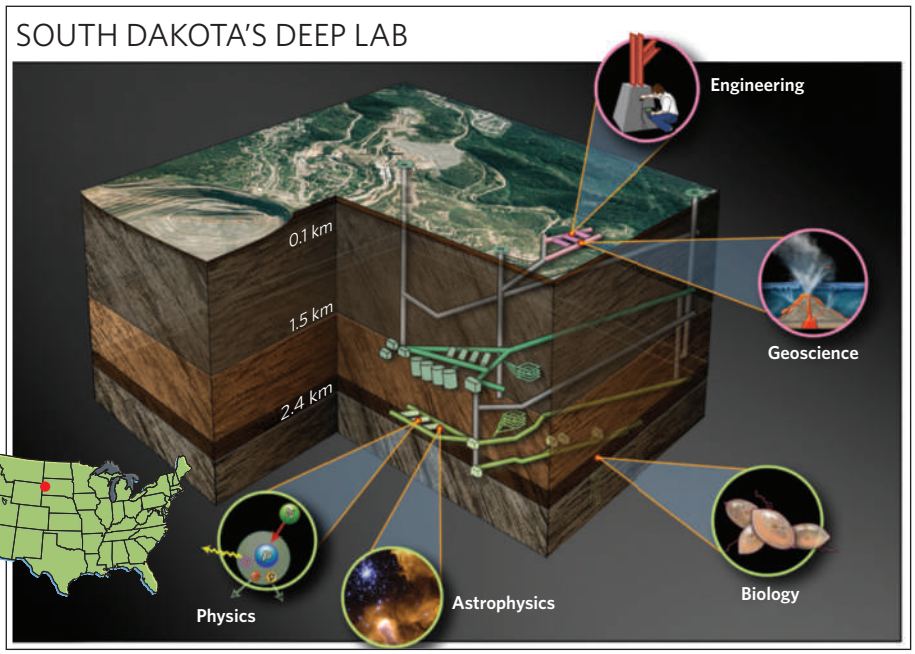
The United States' flagship underground laboratory is running into challenges over its relations with local Native Americans. **Rex Dalton** reports.

Deep in South Dakota's Black Hills, engineers are halfway through pumping water from a 2.6-kilometre-deep mineshaft near the town of Lead. By 2015, US researchers hope, this watery hole will have dried out and become home to one of the country's biggest science infrastructure projects: the Deep Underground Science and Engineering Laboratory, or DUSEL.

But the US$500-million plan has found one of its most difficult tasks on the surface. It has struggled to meet goals to work with local Native Americans, whose cooperation is vital to keeping the project on track. A federal review this year questioned whether DUSEL would create educational and outreach opportunities for local tribes; if not, it could face lawsuits, delays or other major problems.

Project leaders at the National Science Foundation (NSF), the Department of Energy and within South Dakota envision DUSEL as a major facility spinning off jobs for the local community. It landed in the state in 2007 after a hard-fought nationwide competition, lured in part by $120 million from South Dakota, including $70 million from local philanthropist Denny Sanford.

DUSEL will be built within North America's deepest gold mine, Homestake, where physicist Raymond Davis built his pioneering neutrino detector in the 1960s, 1.5 kilometres down. Today, DUSEL leaders plan to install one suite of scientific instruments near the site of that original neutrino work, and to add a second facility 2.4 kilometres down. The idea is to use the overlying rock to block cosmic rays, which can interfere with astrophysics experiments. The facility will join other such laboratories around the world, including the 2.1-kilometre-deep Sudbury Neutrino Observatory in Ontario, Canada, and Italy's 1.4-kilometre-deep Gran Sasso National Laboratory.



Z. DERETSKY/NSF

At DUSEL, the planned experiments include beaming neutrinos to Homestake from the Fermi National Accelerator Laboratory, 1,300 kilometres away in Batavia, Illinois. Detectors in the mine will look for evidence of neutrino oscillation, in which neutrinos change their 'flavour'. Other experiments would hunt for dark matter, study water flow at depth and investigate buried microbial life.

For instance, Tullis Onstott, a biogeologist at

Sex scandal allegations surface at South Dakota school

Allegations of the abuse of women students, cover-ups and retaliations have quietly simmered for years at the South Dakota School of Mines and Technology in Rapid City. The alleged offences include offering grades for sex, physical assault and videotaping of sexual sessions. Now the old allegations, and some new ones, raise questions about how the school will handle human-resource issues under its US$1-million-a-year subcontract for the Deep Underground Science and Engineering Laboratory (DUSEL).

In 2006 , geology professor Gale Bishop sent a dossier to school executives describing more than a decade of alleged personnel violations. The executives have said the issues were investigated and found to be without base. But interviews with students and faculty members suggest the complaints weren't fully probed.

Bishop, now retired, sent the same dossier this March to the school's new president, Robert Wharton. Bishop says he did so because he was concerned that school palaeontologist Gerald Grellet-Tinner was being penalized for, among other issues, trying to help women students who had claimed harassment. Grellet-Tinner was notified in March that his contract would not be renewed.

Grellet-Tinner says he experienced professional and personal disagreements with long-time school palaeontologist James Martin, who is named in the 2006 dossier as allegedly being involved in sexual-harassment cases.

The dossier alleges Martin made 160 videotapes of himself in sexual sessions that included a woman from his field school and Julie Smoragiewicz, vice-president for public relations and admissions. The tape cache was discovered in 2000 by one of the women who was taped, the dossier says; three of the women then destroyed them. *Nature* confirmed these events with four former students or staff with direct knowledge.

Martin and Smoragiewicz declined to be interviewed. Former Earth sciences dean William Roggenthen, now the school's co-principal DUSEL investigator, says he learned of the allegations about four years ago, and required that staff and students be trained on sexual-harassment issues.

Roggenthen says he is unaware of any current issues. "It is my direct responsibility to make sure such inequities don't leak into the DUSEL project," he says. **R.D.**

Princeton University in New Jersey, is looking to DUSEL to understand how microbes function in rock fracture zones. "It is impossible to do this anywhere else; you have to get down there and do the experiments," he says.

Next month, the NSF will award up to $15 million in three-year grants to develop experimental designs for DUSEL projects. In 2011, the agency will ask the National Science Board for preliminary design approval for the full thing. Department of Energy officials say they will seek at least $200 million for complementary facilities.

But local Native American tribes are wary. Long ill-treated by the federal government, who seized the land for its gold more than a century ago and then polluted it with mine run-off, they are cautious about the new influx of government scientists. Physicists visited local powwows to stamp out rumours about Homestake being turned into a nuclear waste dump. Passion and bitterness still run strong, even among Native American scientists.

National project leaders have found themselves entangled in this history. "I'm very much in the learning mode," says Kevin Lesko, a physicist at Lawrence Berkeley National Laboratory (LBNL) in California and the project's principal investigator. "We really want to understand the tribal colleges' perspective."

BHCAIH

"Everything is going a lot slower than I had hoped."
— Jeffrey Henderson

Last summer, LBNL scientists started a cultural advisory committee — the first significant outreach in five years to Native Americans — which has included four Native American scientists and engineers. "We understand we are working in lands that are sacred to Native Americans," says retired California physicist George Campbell, who chairs the committee.

But Campbell has found himself facing a cultural chasm between the Native American community and the South Dakota School of Mines and Technology in Rapid City, which the LBNL has subcontracted to act as the main partner in DUSEL. The school, founded to train engineers shortly after Homestake opened in 1877, has only about 40 Native Americans among its 2,000 students, although 11% of South Dakota's population are Native Americans.

"There has been a very chequered past with the School of Mines," says James Rattling Leaf, an environmental scientist at Sinte Gleska University on the Rosebud Indian Reservation. "Building scientific relationships with them is very tough." By contrast, he says, LBNL scientists have been much more inclusive.

Late last year, the NSF commissioned a peer-review panel to examine DUSEL education and outreach to Native Americans. But project officials didn't share the full results with the cultural advisory committee's Native American members until *Nature* began enquiring. The document says that more Native American representation was needed for the programme to "deliver its true promise"; it went on to say that a real partnership must be created, and "DUSEL should fully integrate the cultural advisory committee" into planning.

Jon Kotcher, who manages the project for the NSF, says the chance to work with Native Americans is an "exciting opportunity". But DUSEL is still in its early stages, and negotiations take time.

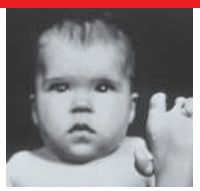
"Everything is going a lot slower than I had hoped," says Jeffrey Henderson, a physician from the Lakota tribe who sits on the cultural advisory committee and is director of the Black Hills Center for American Indian Health in Rapid City. For instance, last spring the South Dakota Science and Technology Authority, which oversees the planned surface laboratory at DUSEL, announced a request for unfunded research proposals. Henderson submitted a proposal to study miners' health, but says he did not hear back. The authority declined to comment.

At a March cultural advisory meeting, project officials gave a presentation on improving Native American educational opportunities through DUSEL. But there was no mention that the School of Mines was no longer working with Rattling Leaf on this type of outreach.

Geological engineer William Roggenthen, the School of Mines' co-principal investigator for DUSEL, says he thinks the school has "really good connections" with some tribal schools. "My institution is committed to doing more," he says.

To learn more about this effort, *Nature* sought to interview other scientists and staff, but they declined to talk, for fear, they said, of retaliation from administrators (see 'Sex scandal allegations surface at South Dakota school').

Campbell tried to smooth over relationships by seeking a tribal blessing for DUSEL. Henderson requested one from a highly respected traditional healer from the nearby Pine Ridge Reservation. The healer declined, saying that despite its good intentions, DUSEL is another exploitation of Mother Earth. ■



G. T. SHARRER, NAT. MUS. AM. HIST.

EXPLAINING THALIDOMIDE'S LEGACY
Drug's effects on embryonic blood-vessel growth may be source of malformed limbs.
www.nature.com/news

Austria quits CERN after 50 years

Austria has announced that it will withdraw from CERN, Europe's premier high-energy physics laboratory, located near Geneva in Switzerland. The announcement — just months before the restart of the Large Hadron Collider (LHC), the world's most powerful particle accelerator — has left Austrian physicists stunned.

"It is a black day for Austrian science," says Christian Fabjan, who heads the Institute for High Energy Physics at the Austrian Academy of Sciences in Vienna. Fabjan says that he was "totally shocked" by the announcement, which was made on 8 May by Johannes Hahn, the science minister and a member of the conservative Austrian People's Party (ÖVP).

Only two other nations have withdrawn from CERN in its 55-year history: Yugoslavia pulled out of the lab in 1961, and Spain left in 1969, only to rejoin in 1983.

Austria joined CERN in 1959, one of the first nations to do so. Two of the laboratory's directors, Willibald Jentschke and Victor Frederick Weisskopf, have been Austrian-born, and the country has 170 scientists working on the LHC and its two largest experiments, ATLAS and the compact muon solenoid. Under the terms of the withdrawal, Austria's participation would end in 2010.

"Nobody is happy about the decision. We would have loved to stay in CERN," says Nikola Donig, a spokesman for the Austrian ministry of science. But, he adds, "budgets are tight". Austria's budget, completed this April, actually increases funding for science, he says. But private funding for basic research has dropped off drastically since the start of the economic downturn.

"Nobody is happy about the decision."

The government will use its contribution to CERN — roughly €17 million (US$23 million) per year, or 2% of the laboratory's budget — to make up some of that shortfall and to begin participation in other international collaborations in physics, sociology and biotechnology. Among those are the European Biobanking and Biomolecular Research Resources Infrastructure project, the European X-ray Free Electron Laser near Hamburg, Germany, and the Facility for Antiproton and Ion Research in Darmstadt, Germany.

Donig says the decision is about getting the greatest return for the government's money. "We want to fund fields where we can have more impact for businesses and universities," he says.

On 11 May, Rolf-Dieter Heuer, CERN's director-general, held what he described as a "constructive" meeting with Hahn. "I think we can still negotiate," Heuer says. He hopes that officials from CERN and the Austrian government can meet in the coming weeks to discuss ways to continue the nation's participation.

The decision still has to be approved by Austria's government, parliament and president. ■

Geoff Brumfiel

Social scientists join synthetic-biology centre

The United Kingdom's first publicly-funded centre devoted to synthetic biology, which opened on 12 May, is hoping to pre-empt public concerns about the field by integrating social scientists into its research team.

Graduate students and staff at the Centre for Synthetic Biology and Innovation at Imperial College London will be trained to consider the social and ethical implications of their research. Sociologists on the staff will also work with government and industry to develop a suitable framework for regulating the products of synthetic biology, and for making intellectual-property claims.

"If the Imperial centre works, they're going to be setting the standard for this," says Pam Silver, a synthetic-biology researcher from Harvard University.

The Engineering and Physical Sciences Research Council is providing the centre with £8 million (US$11 million) over the next 5 years.

For a longer version of this story, see http://tinyurl.com/pc4n9n.

South Africa's cabinet a mixed bag for science

Academics have welcomed the appointment of Naledi Pandor as science and technology minister in South Africa's new cabinet, announced on 10 May by President Jacob Zuma.

Pandor previously headed the education department, but some say she may lack the scientific know-how of her predecessor, mathematician Mosibudi Mangena.

Many researchers had hoped Zuma would retain Barbara Hogan, the respected health minister whose appointment in September 2008 signalled a reversal of the government's denial that HIV causes AIDS.

Instead, Hogan was transferred to public enterprises, and replaced by Aaron Motsaledi, a little-known physician.

For a longer version of this story, see http://tinyurl.com/qczrtg.

GALLO IMAGES / BEELD

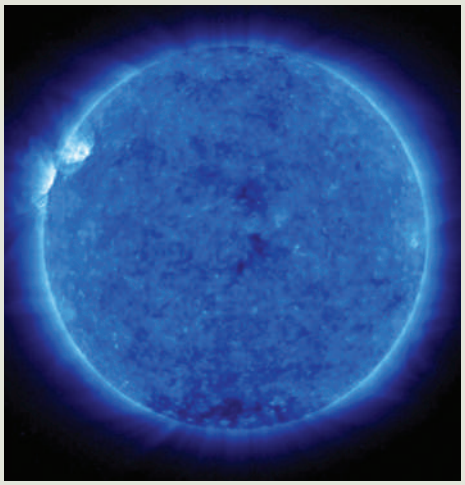
Naledi Pandor has been appointed science minister.

Quiet Sun enters new sunspot cycle

After a prolonged lull in activity, sunspots, and their associated solar storms, are on the rise again.

According to a panel of scientists led by the US Space Weather Prediction Center, part of the National Oceanic and Atmospheric Administration, a minimum in sunspot activity was passed in December 2008. In a consensus forecast on 8 May, the researchers said a new cycle of solar storms would peak in May 2013. But judging by the historical record, the recent persistence of a quiet Sun suggests that sunspot activity at this peak will be the weakest since the solar maximum of 1928.

NASA's STEREO mission spotted large regions of magnetic activity (white spots on image) on the Sun this month.



NASA

University fined after safety-failure lab death

The University of California, Los Angeles (UCLA), has been hit with a fine for multiple safety violations, following the death of a chemistry researcher in a lab fire.

The California Division of Occupational Safety and Health fined the university nearly US$32,000 on 4 May, after the death of Sheharbano 'Sheri' Sangji. Sangji, 23, died on 16 January after being critically burned on 29 December 2008 in the Molecular Science Building.

The university was criticized for failing to train personnel in the use of dangerous chemicals, for not requiring the wearing of protective clothing and for not correcting deficient lab practices identified in an inspection last October.

Gene Block, UCLA chancellor, said in a statement that the university has embarked on a campus-wide review of laboratory safety and practices.

For a longer version of this story, see http://tinyurl.com/cj9mp8.

Human space-flight review in US budget proposals

US President Barack Obama will convene a panel of experts to evaluate the future of NASA's human space-flight programme.

The review will look at whether the International Space Station should be used past 2016, and at the architecture of Constellation, the system of rockets and capsules that will take astronauts back to the Moon. A report is expected by August.

The announcement came as part of the 7 May unveiling of Obama's full budget request for fiscal year 2010. Most science agencies received their top-line funding requests in March, but last week the National Institutes of Health (NIH) got its number for the first time: $30.8 billion, a $443-million increase (or 1.4%) over last year.

Some researchers say the rise is not enough to maintain the momentum they hope to achieve with the $10.4-billion boost for the NIH granted in February as part of the economic stimulus package.

More than 40% ($181 million) of the new money requested by Obama would go to the National Cancer Institute. Across the agency, total spending on cancer research would grow by $268 million, to 5% above 2009 levels.

For more on the US budget, see http://tiny.cc/Zdrjx.

Japan to pay firms to relieve postdoc glut

Japan's science and education ministry has announced a ¥500-million (US$5-million) plan to pay companies to hire postdoctoral students.

The scheme aims to deal with a glut of unemployed postdocs in the nation. The number of academic posts available to them has shrunk since the 1990s, as a result of government streamlining in the university system (see *Nature* 449, 1084–1085; 2007).

By February 2009, 17,827 unemployed postdocs had registered with the Japan Information Career Network (JREC-IN), a website hosted by the Japan Science and Technology Agency (JST) that lists science-related jobs.

Industry traditionally recruits undergraduates, but the JST plans to provide grants to around 100 companies that hire postdocs, mainly through the JREC-IN. The grants would be financed through a supplementary budget being discussed in the current parliamentary session, which is expected to end in June.

Yingxiu, the town nearest the epicentre.

J. QIU

THE SLEEPING DRAGON

The great Sichuan earthquake of 12 May 2008 caught Earth scientists off guard. A year on, **Alexandra Witze** reports from the shattered towns on how researchers have learned from their failures.

Tucked below towering hillsides in Bailu, in China's Sichuan province, two school buildings face one another across a courtyard. Both are several storeys high, white with cheery light-blue trim. It's a peaceful April day, cool and humid; a rubbish bin shaped like a penguin sits at the side of the courtyard, as if waiting for someone to toss in a candy wrapper. But no one will be feeding the penguin today. That's because a nearly 2-metre-high ridge of buckled and uplifted concrete runs right through the courtyard, a manifestation of the geological faults that spawned the great Sichuan earthquake of 12 May 2008.

Along the third side of the courtyard is a ghost. It is a pile of brick rubble, all that remains of another building that collapsed in the quake. There, geologists are hunting for clues to what happened on that day, digging a 40-metre-deep trench to search for signs of past quakes that emanated from these faults.

These cracks in Earth's crust are deceptive pieces of geology. Both Chinese and Western scientists had mapped them before but failed to recognize their potential. "I was astonished at this quake," says Xu Xiwei, deputy director of the Institute of Geology at the China Earthquake Administration in Beijing. The buildings that collapsed and the landslides and mud flows that buried towns combined to kill at least 70,000 people and cause widespread ecological damage (see 'Pandas in peril', overleaf) in this rural corner of southwest China.

More so than other quakes, this one has uncovered gaps in earthquake hazard research, both in China and elsewhere. When scientists assess seismic risk, they tend to focus on the faults that move the most and produce large earthquakes often. That strategy pays off with the many quakes that play by the rules. In western Sichuan, however, it turned out to be disastrously wrong.

One year later, researchers are probing the deadly faults in the hope of finding ways to avoid repeating their mistakes. In retrospect, they say, the geology of the Longmen Shan, or Dragon's Gate Mountains, was trying to warn them.

Mountains of trouble

The range marks the line where the 5,000-metre-high Tibetan plateau rams into the low, stable Sichuan plain. The region has the steepest topographical relief in the world, says geologist Clark Burchfiel of the Massachusetts Institute of Technology (MIT) in Cambridge: over a distance of just 50 kilometres as the crow flies, surface elevation changes by more than 4 kilometres. The Longmen Shan are a world of sloped hillsides cut by dramatic river valleys, the ideal place for quakes to trigger enormous landslides.

That kind of topography does not persist without active geological forces at work, continually building the steep mountain belt. In the

late 1980s, when Burchfiel and his colleagues began mapping the area, they were convinced they would find evidence for large ground movement along the Longmen Shan: perhaps 10 millimetres per year of 'shortening', in which the plateau and plain converge and push up the mountain range.

But years of walking the faults unearthed no evidence for this amount of shortening in the recent geological past. By mapping rock formations, the team found evidence of just 1–2 millimetres of movement per year, instead of the 10 they were expecting. "At that rate, you don't expect to have a mountain range that high," Burchfiel says. Nonetheless, he couldn't deny what the rocks were saying, so eventually he published a major geological overview of the region, supposing that no one would believe the low rates of shortening. Then Burchfiel moved on to map other nearby areas.

"I don't think there was a reason to say there would have been a major quake here." — Leigh Royden

Over time, however, studies have confirmed his conclusions. Researchers measured ground motion in the area using Global Positioning System receivers and found low rates of slip across the Longmen Shan, confirming the 1–2 millimetres per year suggested by Burchfiel[2].

To a geologist, that rate seems relatively benign, because faults store up potential earthquake energy in proportion to the speed of the regional crustal motion. Take two spots on either side of a mountain range, for example. If one is moving quickly in relation to the other, the stress on rocks in between will build up quickly — stress that has to be released by rock movement along a fault. In most cases, that movement is not steady but happens only infrequently, when the stress grows great enough to overcome the friction between rocks on either side of the fault. That sudden release is the earthquake.

In the Sichuan quake, which measured 7.9 on the moment magnitude scale, there was nearly 5 metres of slip along the Beichuan fault, the biggest of the faults that ruptured last year (see map). Given how slowly stress accumulates in the region, rough calculations suggest that quakes of that scale should occur very infrequently, about every 2,000 to 10,000 years[3].

Large shocks in the past will have left their marks in local geology. But the record is hard to read in the Longmen Shan: heavy rains and high erosion rates have obscured much of the evidence, says Alexander Densmore, a geologist at Durham University, UK, who has mapped faults in the area. "There aren't that many places that you can really see the past history," he says. Most of the recent known quakes along the Beichuan fault have been much smaller than the 2008 quake, including one magnitude-6.2 quake in 1958 and another in 1970, says Chen Zhiliang, a geologist at the Chengdu Institute of Geology and Mineral Resources. There is no archaeological evidence that the town of Beichuan itself has ever been destroyed by a quake since it was founded some 1,500 years ago.

So few thought that the Longmen Shan posed a major seismic hazard. "I don't think there was a reason to say there would have been a major quake here," says Leigh Royden, a geophysicist at MIT who has modelled the region's tectonics.

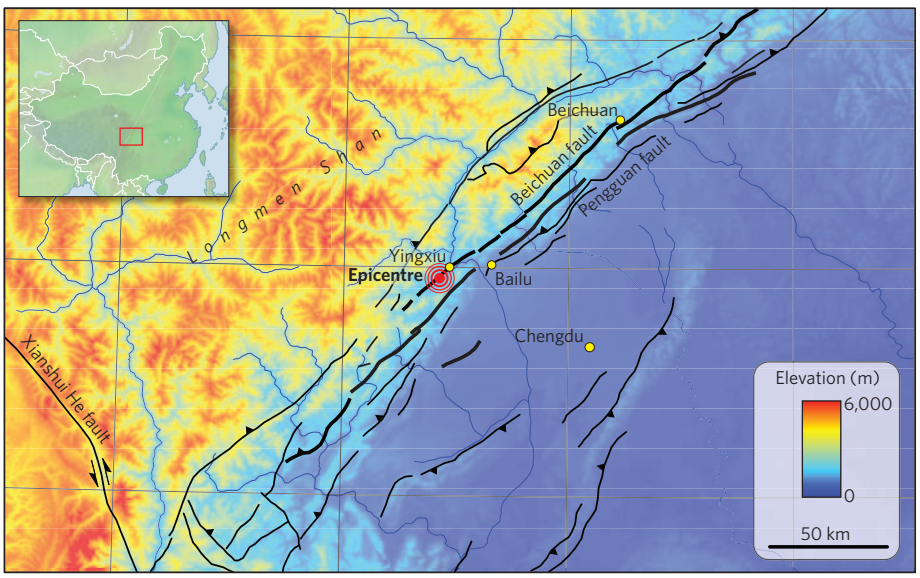
A. WITZE

A fault ripped right through a school in Bailu.

In hindsight, it's easy to see the danger of dismissing the quake potential of the Longmen Shan. Just because something happens rarely does not mean it will never happen. It should have been obvious that the faults along that range were sleeping dragons that would awake some time. But researchers have only so much time and money to spend on seismic-risk assessments, and they therefore focus on areas that are known to have major quakes every few hundred years — not ones that might stay quiet for 5,000 years.

For example, rather than worry too much about the Beichuan fault, Chinese geologists had focused on a pair of far more active fault zones to the west: the Anning He and the Xianshui He faults, both of which slip at rates of up to 10 millimetres per year. The China Earthquake Administration has spent most of its monitoring efforts on these active faults, including deploying nearly 300 broadband seismometers — ones that capture a wide range of vibrational frequencies — in the world's densest array to map the underlying crust. When the Beichuan fault broke instead, seismologists scrambled to refocus on the Longmen Shan. Some are now looking into whether a new reservoir nearby triggered the quake (see 'The reservoir link', overleaf).

The question now is what the Sichuan quake tells geologists about future seismic risk. Some say that more attention should be paid to regions with steep topographical relief, even if



they have minimal ground movement. Royden points to an analogous region in Canada's Northwest Territories, but few people live there, so it is unlikely to become a priority for research. Within China and in other densely populated regions, there are few obvious analogues, although researchers will surely be taking a fresh look at mountainous zones.

Beyond being deceptively lethargic, the Beichuan fault caught Earth scientists off guard last year in another way. From the surface, the fault appears to be divided into relatively short segments that were assumed to move separately in relatively small earthquakes. "We traditionally tend to look at individual fault segments and say those are the maximum size of the earthquake," says John Shaw, a geologist at Harvard University. But if the segments can connect, "the magnitude of those earthquakes is much greater than anticipated".

That is what happened last year. The Beichuan fault ruptured across several segments totalling 240 kilometres, while a secondary fault to its southeast, the Pengguan fault, broke for 72 kilometres. The segments apparently connect at depth, allowing the quake to grow larger than would have been expected. Chinese geologists are now beginning to map in detail the faults that connect with the Beichuan fault.

The danger that remains is another concern. Because the Beichuan fault broke almost entirely to the northeast of its epicentre, some scientists wonder whether the segment that runs towards the southwest is ready to go. Nearby faults may also pose a risk. One study suggests the Beichuan quake increased stress, among other places, on the Xianshui He fault, and on other faults near the city of Ya'an and southeast of Chengdu, the capital of Sichuan[4]. Another study proposes that the chance of a magnitude-7 or greater quake in the area during the next decade is now 8–12%, higher than it was before the 2008 quake[5].

Pandas in peril

"The mountain literally exploded. Boulders were flying in the air, along with earth and leaves," says Deng Linhua, recalling the Sichuan earthquake on 12 May 2008. "It was horrifying." Deng is a vet at the Wolong National Nature Reserve, China's most famous sanctuary for giant pandas and other rare animal species, which sprawls across 200,000 hectares of rugged terrain just a few kilometres from the quake's epicentre.

The earthquake set off landslides, debris flow and avalanches, causing serious ecological damage to 15% of the panda habitat in Wolong. In particular, the reserve's bamboo forests were badly hit, threatening the future of its 143 wild pandas, about one-tenth of the total wild population in China.

The quake also destroyed the China Conservation and Research Centre for the Giant Panda, which boasts the world's largest breeding and research programme, established in 1980 in Wolong by the Chinese government and the international conservation group now known as WWF. "The pandas were terrified," says Deng Tao, a panda keeper at the centre. "Some ran away, and others climbed up to the top of trees and wouldn't come down."

The centre's staff evacuated most of their 63 captive pandas about 300 kilometres by road to another panda base in the area of Ya'an, south of Wolong, and to zoos that were able to house pandas (see photo). One panda is still missing, and one mother of five was killed in her enclosure by a landslide.

Six young pandas remain at the centre in Wolong. One year on, they are happily savouring fresh bamboo in their new, temporary enclosures. "They have mostly recovered from the trauma, but are very sensitive to loud noises," says Deng Tao.

Wolong's reconstruction is also under way: the conservation and research centre will be relocated to an open valley near Gengda, more than 10 kilometres northeast of the current site. "The new site will be less vulnerable to geological hazards," says Wang Lun, the reserve's vice-director.

A total fund of 200 million yuan (US$29 million) has been earmarked for this project and other reconstruction efforts at Wolong, which are expected to take three years to complete. **Jane Qiu**

WOLONG NATIONAL NATURE RESERVE

How it hit

The biggest city in this threatened zone is Chengdu, now teeming with 10 million people. Constant traffic jams and high demand make it near impossible to hail a taxi during working hours. Young professionals who have relocated from Beijing or Shanghai to enjoy a more laid-back lifestyle thread their way through the crush on electric bikes. Ethnic Tibetans, part of the diverse mix in southwest China, find themselves shouldered out of the boom, and many end up as beggars on the pavement.

"People often forget to account for disaster prevention in reconstruction." — Cui Peng

Chengdu is also home to the province's leading earthquake scientists, for whom the 12 May quake — referred to as '5/12' for short, like '9/11' in the United States — occurred practically in their backyards.

In his tidy office in Chengdu, with a Chinese-language copy of *On the Origin of Species* at hand and a picture of Albert Einstein looking on, Chen recalls what it was like at 2:28 p.m. on 12 May. The office began shaking with the strongest tremors he had felt in more than 40 years in the city. Staff members evacuated the building; people poured into the streets as bricks rained down. Chen tried to call his son, but phone lines were dead, so he rushed to the nearby primary school to find his granddaughter. Then he came back to his office building, which was constructed to some of the highest quake-protection standards in the city, and within two days had posted online a history of quakes in the Longmen Shan area.

Across town, when the quake hit, geodesist Du Fang hid under the sturdy wooden table in her office at the Sichuan seismological bureau. Du, the deputy director for earthquake prediction, says she had no idea the quake was coming. Although there are anecdotal reports of toads pouring into Sichuan streets as indicators in the days before the quake, scientific data justify her. Seismometers along the Beichuan fault recorded no increase in tremors that might have presaged the quake, although one station 640 kilometres south of Beichuan recorded changes that some claim were a warning.

In many ways, the Chinese government is still struggling with the aftermath of the disaster. Praised initially for its quick response in sending emergency crews into the affected areas, the government soon faced angry parents asking why so many schools had collapsed. Bitterness lingers. In Yingxiu, the town closest to the epicentre, where as much as 80% of the population died in the quake, rows of temporary housing crowd up against the ruins of Xuankou Middle School where 55 people, including 43 students, were crushed to death.

The government is rebuilding at breakneck pace. Above each cluster of temporary shelters

rises an optimistic billboard showing gleaming plans of new houses to be built. Some are already done: fresh paint and new concrete rise from the recently cleaned-up hillsides, with red characters for good wishes inscribed over the brand-new doorways.

New houses along the Longmen Shan are supposed to be able to withstand a magnitude-8 quake; previously, building regulations in Chengdu required construction to withstand only a magnitude-7 quake, which has one-tenth the intensity of shaking. In many places, however, reconstruction is taking place so quickly that no one is confident that building codes are being followed. Villagers bring hand-carts to the landslides that once blocked the road and haul away rocks to break them and use them to start building homes afresh. Piles of brick — one of the worst construction materials for a quake-prone zone — dot the sides of main roads, waiting to be mortared together into new homes. Lorries piled with construction materials cause hours-long traffic queues along the narrow roads that thread through the mountain valleys.

Even as construction cranes rise from town centres, the landslide-scarred mountains above tower ominously. More than one-fifth of the people who died in the quake were killed by landslides or mud flows, says Cui Peng, a geomorphologist at the Institute of Mountain Hazards and Environment in Chengdu. Precise numbers are hard to tally — the affected area sprawls over 130,000 square kilometres

A. WITZE

Delving into earthquake history at Bailu.

in 51 counties — but estimates suggest that at least 50,000 landslides occurred, perhaps as many as 100,000 or more[6]. One, in Wangjiayan, killed 1,600 people. Another, at Beichuan High School, buried 400 students. Elsewhere, landslides did not kill directly but dammed rivers, creating more than two dozen major 'quake lakes' that threatened residents downstream.

The danger of landslides, Cui warns, will be even more acute this rainy season, which begins late this month. The quake destabilized a number of slopes in the area, making them particularly prone to failure after rain. Last September, for instance, heavy rains sent a mud torrent sweeping into the empty centre of Beichuan, already devastated by the earthquake months earlier. The problem is exacerbated by large-scale damage to the landscape from mining practices that have carved out hillsides, and from deforestation that has stripped the slopes of their protective trees.

If people rebuild houses in places that are prone to landslides, Cui notes, constructing them to withstand quakes won't help. "People often forget to account for disaster prevention in reconstruction," he says. His team at the institute, which is part of the Chinese Academy of Sciences, has made detailed recommendations to the government to highlight areas that should avoid rebuilding. Meanwhile, new houses are springing up informally in the villages that dot the Longmen Shan — one by one, and probably not in government-approved areas.

A flood of data

Amid the disheartening news, however, scientists say the data from the earthquake itself will illuminate the region's geology at a more fundamental level. Those data exist because in recent years the Chinese government has spent a lot of money on new equipment to try to make its Earth sciences competitive in the world arena.

The reservoir link

A. WITZE

Just 15 kilometres from the epicentre of the 2008 Sichuan earthquake, the concrete wall of the Zipingpu dam has tamed the once-turbulent Min River to form a placid lake. Controversy belies the calmness of its waters, however.

Christian Klose, a geologist at Columbia University in New York City, argues that filling the reservoir, which began about 2 years before the quake and eventually impounded 300 million tonnes of water behind the dam, would have changed the stresses on the regional fault system and may have pushed the Beichuan fault to rupture in the killer quake.

Klose has submitted his work for publication in a peer-reviewed journal and presented it in December at an American Geophysical Union meeting in San Francisco. Researchers who heard the talk, or have spoken to him about the work, remain largely unconvinced, however. "That quake — there's no way it was induced by a small reservoir," says David Simpson, president of the Incorporated Research Institutions for Seismology in Washington DC.

Simpson acknowledges that reservoirs have been linked to much smaller earthquakes in the past; perhaps the best-known example is the Koyna dam in India, the construction of which helped to trigger a magnitude-6.5 earthquake in 1967 that killed more than 180 people.

For this reason it is common to install seismic sensors around newly built reservoirs, and the Chinese government did just that around the Zipingpu reservoir. The sensors monitored seismic activity as the water rose, then lowered, then was filled to the top.

Those data, say various Chinese sources, do not indicate a significant increase in seismicity before the great earthquake, a signal that might have been expected if adding water had changed stress on the fault system.

Others argue that, although reservoirs might lead to smaller earthquakes near the surface, they could not trigger a magnitude-7.9 quake whose focus was 19 kilometres underground.

More intriguing is the question of whether the construction of the reservoir might have merely prompted a major earthquake that was due to happen anyway. Walter Mooney, a geologist at the US Geological Survey in Menlo Park, California, says that possibility is little relief to those who lost loved ones in the disaster. "Even if the reservoir was slightly related to it or advanced the time for when an earthquake was ready to go," he says, "that's still not socially acceptable."

China is not formally investigating the proposed link, although many Chinese geologists say privately that more work is warranted. **A. W.**

A crown jewel of the government's programme is the array of nearly 300 broadband seismometer stations, which was deployed in western Sichuan by Liu Qiyuan of the China Earthquake Administration and his team. The envy of Western scientists, the array boasts the densest arrangement of seismometers of any large network around the world: it has yielded more than 7 terabytes of data so far. Rob van der Hilst, a geophysicist at MIT who set out an earlier 25-station array in the same region, calls the Chinese network "an enormous tour de force". First deployed in October 2006 and spaced 5–30 kilometres apart, the solar-powered stations cover 370,000 square kilometres of mountainous terrain; someone visits each station every four months to collect the data. Originally funded with 60 million yuan (US$9 million) from the ministry of science and technology and more than 8 million yuan from the provincial government, Liu now scrapes together 1.8 million yuan per year to keep the network operating.

Last May, the great quake knocked out three of the array's stations; one was squashed under a massive boulder. But the data recorded at the time by nearby stations are yielding an unprecedented glimpse into the crust of western Sichuan; a major quake has never been captured in such detail by a network like this. "It's a very rare opportunity in the world," says Liu. "This quake should play an important role in seismological history."

Preliminary data suggest that there is a major change in the geology roughly 20 kilometres below the surface, where relatively brittle material gives way to deeper, softer rock through which seismic waves travel more slowly. This could help explain, Liu says, why the quake and all its aftershocks occurred in the upper 20 kilometres of crust.

"This quake should play an important role in seismological history." — Liu Qiyuan

Liu is now collaborating with van der Hilst and Michel Campillo of Joseph Fourier University in Grenoble, France, to run the data through new seismic analytical techniques[7]. He is also working with scientists from Taiwan, who are interested in probing any possible analogies with Taiwan's 1999 Chi-chi earthquake.

Liu originally set up the network to monitor what had seemed the biggest threat in the region: the Anning He and Xianshui He faults. After the 2008 quake, however, Liu shifted some of his stations to the north and east, onto the Beichuan fault. The array will remain in place there for a year, after which most of the stations will be moved to other areas, having collected the data he wants.

A. WITZE

The cemetery at Yingxiu commemorates a day when most of the town's citizens died.

Meanwhile, other researchers are trying different ways to investigate the geological history of the Longmen Shan. In a project spearheaded by the land and resources ministry, a team is drilling four holes along the fault zone to collect continuous rock cores from as deep as 4 kilometres. A pilot hole in the village of Hongkou has passed 650 metres' depth and may have already penetrated the fault zone, says Li Haibing, the project's chief geologist, who is at the Institute of Geology and Geophysics of the Chinese Academy of Sciences in Beijing. Team leaders intend to put seismological instruments down the hole for long-term monitoring.

At the shuttered Bailu school, just off its tortured courtyard, the palaeoseismology trench is getting ever deeper. The pit, hand-dug by workers carrying buckets of dirt on yokes, has already revealed evidence of past tremors. Arcing layers of cinder mark the remains of fires triggered by smaller quakes like those that occurred in 1958 and 1970. The results from this trench, along with studies of the buildings still standing in Bailu, may aid future planning. Xu says that the government may intensify mapping of all the active fault traces in the country, in the hope that more precise knowledge may save lives.

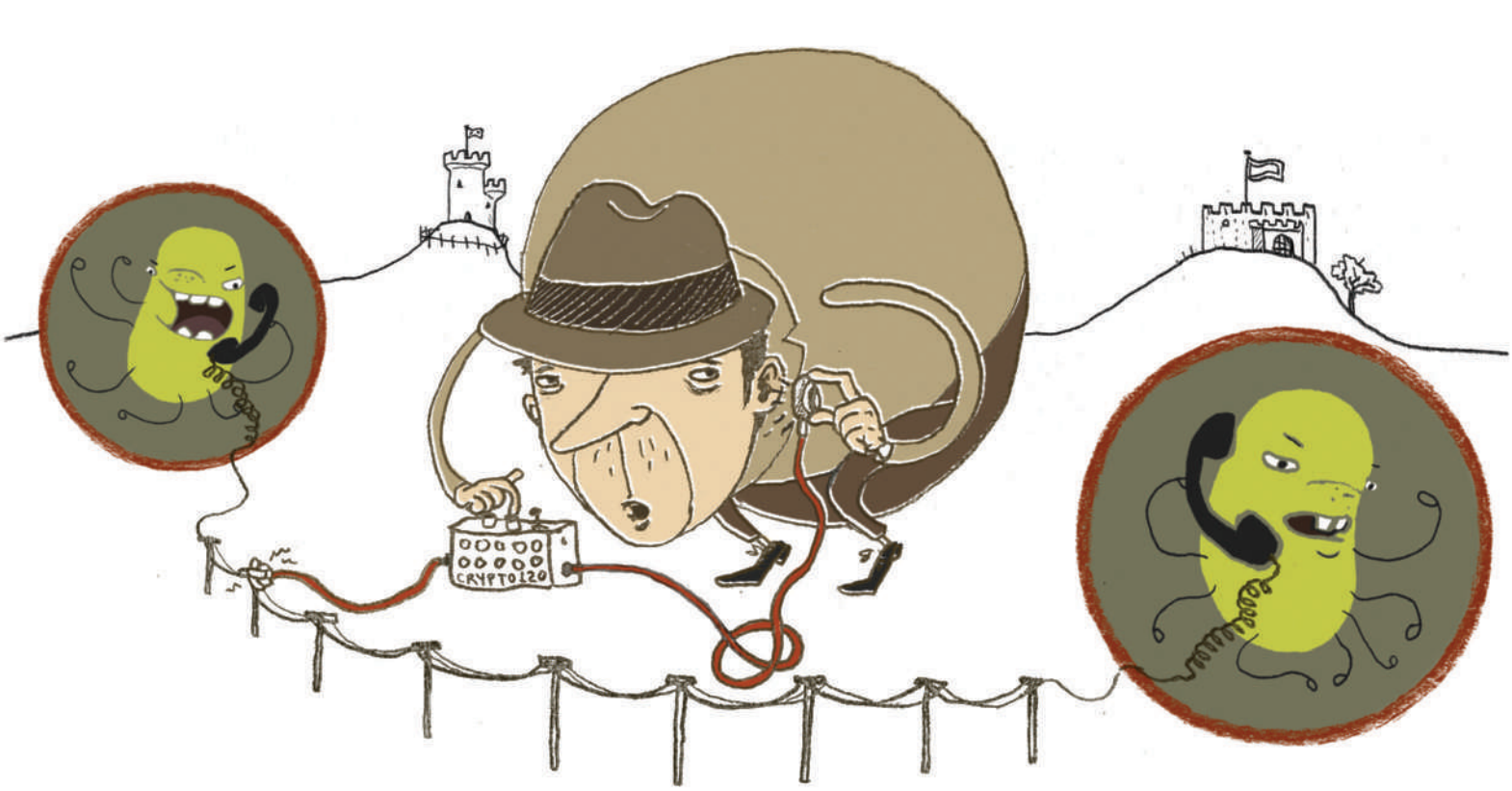
As they look back on the earthquake, Earth scientists in China and around the world say that they remain chastened by their lack of foresight. Although many say they could not have recognized a hazard that rears its head only once every few thousand years, the recent disaster has made researchers rethink their assumptions, especially in areas where geological forces are so evidently at work. In the future, they will be less likely to conclude that areas showing little evidence of movement are safe from large quakes.

That will come as little consolation to the people of Bailu. On a spring day, a group of children swarms over a concrete court in town, shouting and elbowing each other in a game of basketball near the abandoned school. Up above, a caged songbird overlooks the playground for good luck. Rows of vegetable gardens dot the hillsides, fresh green against newly tilled dirt. But the school itself remains closed for good, a memorial park to the victims of the 2008 quake. ■

Alexandra Witze is *Nature*'s chief of correspondents for America. Additional reporting by Jane Qiu, *Nature*'s retained correspondent in Beijing.

1. Burchfiel, B. C., Chen, Z., Liu, Y. & Royden, L. H. *Int. Geol. Rev.* **37,** 661–735 (1995).
2. Zhang, P.-Z. *et al. Geology* **32,** 809–812 (2004).
3. Burchfiel, B. C. *et al. GSA Today* **18,** 4–11 (2008).
4. Parsons, T., Chen, J. & Kirby, E. *Nature* **454,** 509–510 (2008).
5. Toda, S., Lin, J., Meghraoui, M. & Stein, R. S. *Geophys. Res. Lett.* **35,** L17305 (2008).
6. Cui, P. *et al. Nat. Hazards* advance online publication doi:10.1007/s11069-009-9392-1 (2009).
7. Courtland, R. *Nature* **453,** 146–148 (2008).

See Editorial, page 140.



TINKER, BACTERIA, EUKARYOTE, SPY

Bacteria and their hosts may reside in different kingdoms, but that doesn't stop them from intercepting each other's communications. **Asher Mullard** reports.

When Mark Lyte looked up from the podium at the 1992 American Society for Microbiology meeting in New Orleans, Louisiana, he saw two faces — and nearly 400 empty seats. Lyte, a microbiologist at Texas Tech University in Lubbock, ploughed on regardless. He had a lot to say. His experiments had shown that three species of infectious bacterium intercept the human stress-response hormone noradrenaline and use it as a cue to escalate their growth — perhaps explaining why stressed animals are more likely to die of infection despite having boosted their immune responses.

Minutes into Lyte's lecture, one person got up and walked down the long, lonely aisle to the door. Only his loyal technician remained, along with the two people chairing the session. After the thin applause, one of them posed a question: "Why would you ever want to do these experiments?" Researchers already knew that bacteria and humans detected each other's presence through membrane receptors and cell-wall molecules — but no one thought that bacteria were sophisticated enough to eavesdrop on the long-range chemical signals of the organisms that host them.

Nowadays, Lyte's lectures are packed, and a burgeoning field of researchers is studying the chemical crosstalk between bacteria and their hosts. They know that, as Lyte showed, bacteria respond to the chemical signals that their hosts use for internal communications; they've also discovered that infected hosts intercept the signals that bacteria send to each other, apparently to confound their knavish tricks. The bacteria fight back, turning off immune responses. Interkingdom espionage offers all the intrigue, jamming, fakery and subversion that you could find in a good spy thriller.

Some scientists see all this subversion as something to emulate: if body cells can confuse bacterial attackers with this sort of crosstalk, why shouldn't pharmaceutical companies? Hence the search for small molecules or antibodies that could serve as new classes of antibacterials. And on top of — or beneath — these practical opportunities, there are also deeper scientific questions. How did organisms as different as bacteria and their eukaryotic hosts come to understand each other in the first place? And is it possible that the crosstalk underlies mechanisms of cooperation as well as conflict?

For centuries, bacteria were thought to be loners that didn't communicate with one another, let alone with anything else. Bonnie Bassler, who studies bacterial communication at Princeton University in New Jersey, recalls a time when many of her colleagues thought that "bacteria didn't have the genetic power to do anything interesting — they ate, they moved, they divided". But in the 1970s, researchers discovered that *Vibrio fischeri*, bacteria that live in squid, fish and the open ocean, coordinate their bioluminescence by sensing the level of signalling molecules given off by others[1], a system that later came to be called quorum sensing. Signalling that synchronizes bacterial gene expression patterns and coordinates behaviour within a population has now been seen in all sorts of bacteria — and it is used for various purposes, including establishing infection and increasing virulence.

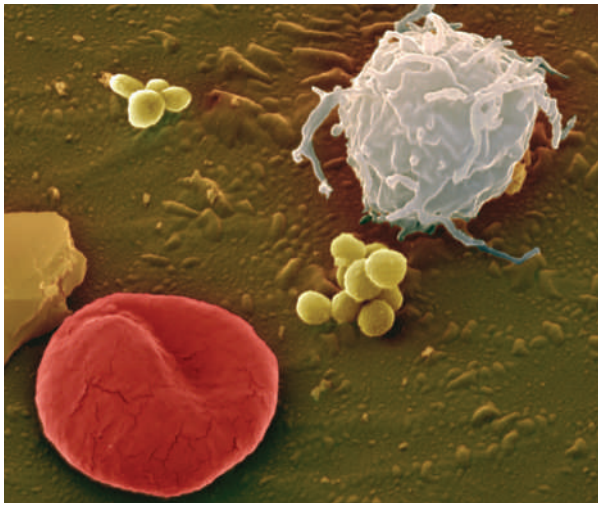
ILLUSTRATIONS BY KAREN CHEUNG

Over the wall

Still fighting against the idea that bacteria were simpletons, Lyte's study[2] in 1992 was one of the first to show that bacteria also detect signalling molecules released by the organisms that they infect. In 2006, microbiologist Vanessa Sperandio, at the University of Texas Southwestern Medical Center at Dallas, and her colleagues showed how intimately the two communication systems could be integrated. Sperandio's team found that QseC, a bacterial receptor that detects a quorum-sensing signal called autoinducer 3 (AI-3), is also activated by the mammalian hormones adrenaline and noradrenaline[3]. Both cause the bacterium *Escherichia coli* to express virulence genes. Sperandio suspects that AI-3 and the human hormones have structural similarities that enable them to bind to

the QseC receptor, and is looking into whether human hormone receptors can also detect AI-3.

Some argue that this crosstalk is not 'signalling' at all — it did not evolve specifically as a means for two willing parties to communicate. But the fact that the same receptor performs this double duty still requires explanation. Some invoke convergent evolution, suggesting that functional requirements led bacteria and their hosts to evolve chemical messengers with some similar characteristics. An alternative possibility — a controversial one, but one to which Sperandio subscribes — is that the same receptor works for both bacterial and eukaryotic signals because eukaryotic cells acquired the genes for cellular communication from bacteria. This was proposed in 2004 by evolutionary biologist and bioinformatician Eugene Koonin, at the National Center for Biotechnology Information in Bethesda, Maryland, and his colleagues[4]. On the basis of the way the genes involved in hormone metabolism are distributed, Koonin argues that cell–cell communication machinery may have been passed from bacteria to eukaryotes on several occasions by lateral gene transfer.



J. BERGER/SPL

Who's listening in? *Staphylococcus* bacteria between two body cells.

Spooking the spooks

However the similarities between the signalling systems arose, they have enabled bacteria and their hosts to indulge in some deception. "What we're looking at is not only espionage, but also hijacking," says Kendra Rumbaugh, who studies interkingdom signalling at Texas Tech University.

Take the microbial messenger C12. This is a quorum-sensing signal that coordinates the expression of virulence genes for *Pseudomonas aeruginosa*, a pathogen that can infect burn injuries or people with supressed immune systems. When Gunnar Kaufman from the Scripps Research Institute in La Jolla, California, started working on it in 2005 he knew it was detected by mammals, and several studies suggested that it helped to trigger inflammation. "But it's the opposite," Kaufman says.

When he and his colleagues treated mice with C12 they found that it actually inhibits the NF-kB signalling pathway[5], which is crucial for immune response; studies on human cells indicated much the same. While it might make sense for the host immune system to listen out for the quorum-sensing signal, in this case the bacteria seem to have evolved the upper hand. "C12 acts as a stealth agent," Kaufman says. "*P. aeruginosa* might use it to shut down the immunity locally so that by the time the host realizes there is something there, it is too late," he says.

Dirty play goes both ways. Plants and algae, for example, are master mimics of bacterial quorum-sensing signals. One of the best-known examples is found in the red alga *Delisea pulchra*, which produces quorum-sensing signal lookalikes called furanones. In 2002, microbiologist Staffan Kjelleberg of the University of New South Wales in Sydney, Australia, and his colleagues showed that furanones jam signalling in *P. aeruginosa* and in *E. coli*, probably either by competing with native quorum-sensing signals or by changing the configuration of the bacterial receptor[6].

Another tactic, discovered in animals, is simply to snatch the messengers off the streets. Hattie Gresham, a microbiologist at the University of New Mexico in Albuquerque, has been studying how hosts handle pathogenic *Staphylococcus aureus* for nearly 15 years. About 25% of people have these bacteria residing permanently in their nose and an estimated 1% live healthily with methicillin-resistant *Staphylococcus aureus* (MRSA). "That means that the host has something that can keep the bacteria in check," says Gresham. She suspected that some component of human blood plasma was interfering with *S. aureus* communication.

In 2008, after painstakingly screening serum samples, Gresham and her team found that component: apolipoprotein B (APOB), a huge lipid-binding protein that helps transport cholesterol in the bloodstream[7]. Gresham found that APOB smothers an *S. aureus* quorum-sensing molecule called autoinducing peptide 1 (AIP1), cutting the line of communication used to coordinate the onset of virulence. Mice chemically or genetically manipulated to lack APOB are more susceptible to MRSA.

In this case at least, says Gresham, both sides benefit: the host prevents an infection from turning pathogenic, and the bacteria are able to live happily in the nose without threat from the host's immune system. Through mutual surveillance and manipulation, the host and the pathogen can "arrive at a détente", says Gresham. If the balance breaks down, because a patient is old, sick or otherwise immunocompromised, then the infection starts escalating out of control. Clinical studies have shown that APOB levels are lower in critically ill patients than in healthy individuals, which Gresham thinks could partly explain why these patients are highly vulnerable to MRSA infection. "Therapeutically, is there a way to lower that risk by giving these patients APOB, or a peptide mimetic of APOB?" she wonders.

Researchers have been trying to manipulate quorum sensing to make antibacterial drugs since the 1990s. They have had little success; a fair few early start-ups based on the idea died. One problem may have been the dearth of knowledge about interactions between bacterial signals and host signals. "It's hard to think about developing a drug that targets quorum sensing without knowing how the host deals with [quorum sensing]," says Gresham. Researchers are wary of blocking a quorum-sensing signal if the host might already be using it to gauge its immune response, or of developing a compound that risks inhibiting both bacterial and host receptors.

Still, many microbiologists and chemists are still hopeful that they can design neat little molecules to artificially stifle or manipulate microbial communication systems more effectively than the systems that have evolved naturally. "This field is kind of like a sandbox for chemists," says Helen Blackwell, a chemist from the University of Wisconsin, Madison. "If we can understand these signals better, and learn what components of these signals

are necessary at the molecular level, then we can tinker with them and start to engage the bacteria in new conversations, and we can try to confuse them."

Working in close collaboration with chemists, pharmacologists and other microbiologists, Sperandio screened 150,000 molecules for inhibitors of the quorum-sensing receptor QseC and identified one, LED209, as a potent, relatively non-toxic small molecule that protects mice from both *Salmonella typhimurium* and *Francisella tularensis*, although not from pathogenic *E. coli*[8]. In 2008, the group won US$6.5 million over 5 years from the US National Institutes of Health to search for LED209 analogues that provide greater protection and lower toxicity. These, they hope, will find use as broad-spectrum therapeutics to protect patients from stubborn infections associated with assisted-breathing apparatus, several of which have receptors much like QseC. "Our idea is to have this in a preclinical form in 5 years," she says.

For Nafsika Georgopapadakou, though, the anti-quorum-sensing approach is ultimately flawed. Georgopapadakou, a consultant in Montreal, Canada, has worked on antimicrobials at several large companies. She says that quorum sensing seems to be important for establishing infections rather than for maintaining them, and so such therapeutics are only likely to be useful as prophylactics that are given before the infection has started. Anti-quorum-sensing approaches don't kill bacteria, she adds, they just lower virulence and increase the odds that antibiotics and the immune system can clear the infection. "If I'm going to sweat making novel compounds, I would much rather kill bugs."

But proponents of anti-quorum-sensing approaches argue that their non-lethal approach is actually advantageous. One of the main failings of current antibiotics is that their efficient killing drives the rapid evolution of drug resistance, says Sperandio. Anti-quorum-sensing strategies, by contrast, could have a much longer shelf life. "If you don't kill the bacteria, you're not speeding up the process of developing resistance that much," she says. To get the full effect, however, they will probably have to be used with other drugs, she adds.

J. VARNEY

"If we can understand the signals better, we can engage the bacteria and try to confuse them." — Helen Blackwell

The spy who loved me

Some researchers are less interested in intervening in bacteria–host communication and more interested in exploring why it happens. "Everybody, including our lab, has focused on pathogenic processes," says Rumbaugh. "Unfortunately, the field might be focusing on the wrong direction." As Rumbaugh sees it, "pathogenesis is the exception". Many microbiologists believe that these cross-kingdom communication systems evolved because they served a beneficial purpose for both sides, by supporting mutually beneficial relationships between bacteria and hosts. "So, what are the real functions of these [interkingdom exchanges]?" Rumbaugh asks.

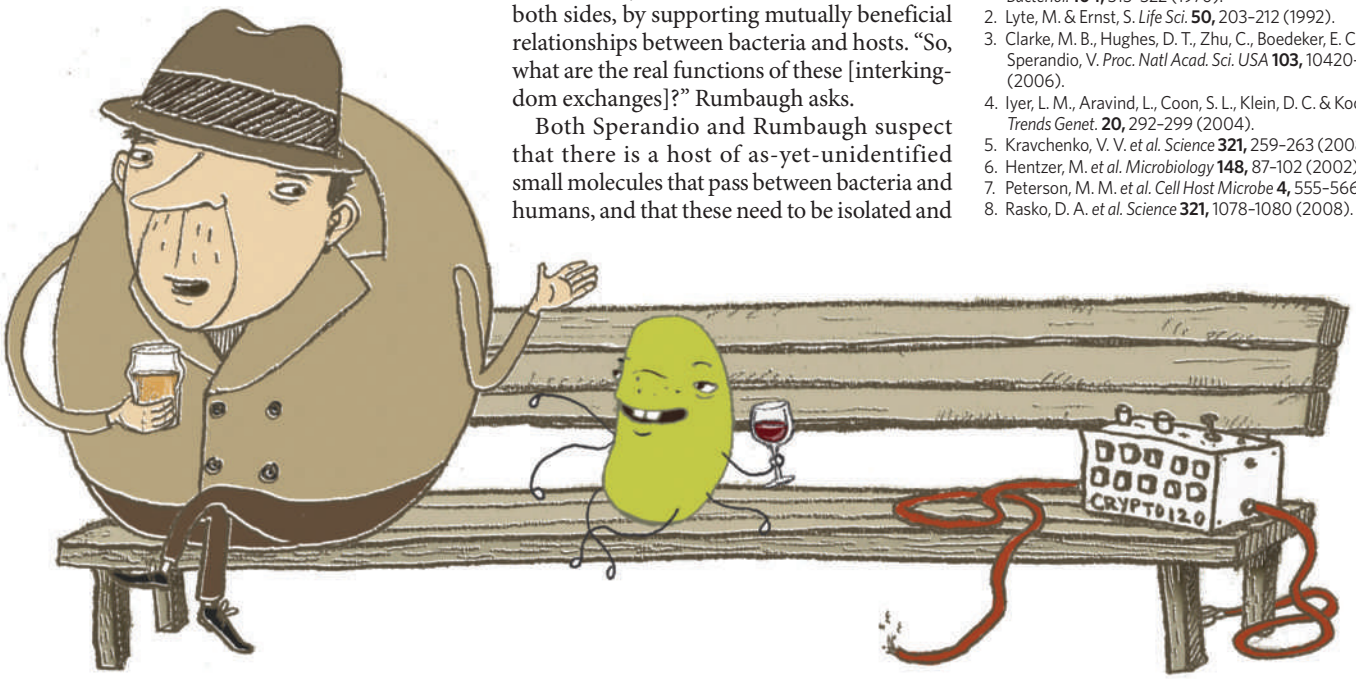
Both Sperandio and Rumbaugh suspect that there is a host of as-yet-unidentified small molecules that pass between bacteria and humans, and that these need to be isolated and catalogued if researchers are going to understand the full scope of interkingdom communication and the purposes that it serves. Practically, this could be tough. Quorum-sensing signals can be hard to distinguish from other soluble chemicals; some of them are only produced in specific environmental conditions that can be near-impossible to reproduce in a culture dish, and some are manufactured in such minute quantities that they are difficult to collect and analyse. The structure of AI-3, for instance, has not yet been resolved for this reason, Sperandio says.

There is an additional obstacle to deciphering friendly bacterial–host communication. Just as vice and espionage tend to capture the fiction market, Rumbaugh says it's easier to win research funding to study duplicity and pathogenicity than it is to study the friendly 'symbiotic' and 'commensal' interactions in which one or both sides benefit.

As for Lyte, he too is pursuing the idea that an extensive and cordial dialogue is going on. He wants to examine whether bacteria use their signals to modulate complex host behaviours and functions, including learning and memory, and vice versa. "Bacteria are conversing with us, and we're conversing with them," says Lyte. The question now is how to record more of the conversation — and work out what is being said. ■

Asher Mullard is a freelance science writer based in London.

1. Nealson, K. H., Platt, T. & Hasting, J. W. *J. Bacteriol.* **104,** 313–322 (1970).
2. Lyte, M. & Ernst, S. *Life Sci.* **50,** 203–212 (1992).
3. Clarke, M. B., Hughes, D. T., Zhu, C., Boedeker, E. C. & Sperandio, V. *Proc. Natl Acad. Sci. USA* **103,** 10420–10425 (2006).
4. Iyer, L. M., Aravind, L., Coon, S. L., Klein, D. C. & Koonin, E. V. *Trends Genet.* **20,** 292–299 (2004).
5. Kravchenko, V. V. *et al. Science* **321,** 259–263 (2008).
6. Hentzer, M. *et al. Microbiology* **148,** 87–102 (2002).
7. Peterson, M. M. *et al. Cell Host Microbe* **4,** 555–566 (2008).
8. Rasko, D. A. *et al. Science* **321,** 1078–1080 (2008).



CORRESPONDENCE

Leading the tributes to editor John Maddox

SIR — In April 1974, some months after I had taken over from John Maddox as editor of *Nature*, I was driving home from the printers with a colleague at four in the morning, having just put the latest issue to bed. News came in over the radio of a coup in Portugal. What would John have done? We agreed that he would have turned the car round and written a new thousand-word Editorial: 'What future for Portuguese science? The coup in Lisbon is, or ought to be, an opportunity for Portuguese scientists ...' We smiled at the thought, but drove on.

This little story exemplifies John's approach to *Nature*. As a one-time journalist, he prized immediacy. He had a formidable list of contacts, and even if he hadn't known any Portuguese scientists, he would still have created a sense of authority.

Until his arrival as editor in 1966, *Nature* had been a worthy journal of record but lacking in flair; it changed rapidly as John brought his journalistic background to bear. 'We wuz robbed' was the title of an Editorial written at the time of the 1966 Football World Cup, proposing a new method for determining the winner. Very different from previous fare, which ran along the lines of 'comment on the progress of Her Majesty's Alkali Inspectorate as described in its 47th Annual Report'.

John gathered around him enthusiasts in the academic world for this new style of journal. He urged us to seek out good scientific papers and gave us free rein to hold forth in Editorials. We were awed by his restless energy in generating thousands of words.

John was immensely active. He took on broader responsibilities within Macmillan; he launched the weekly *Nature New Biology* and *Nature Physical Science*; he spoke regularly on the radio; he challenged environmentalists' excesses and wrote a book, *Doomsday Syndrome* (Macmillan, 1972). That year, he founded Maddox Editorial Ltd, which went on to publish a European journal. The result of all this was that *Nature* received less than his full-time attention and began to fray at the edges. In 1973, Macmillan and John parted company.

Shortly before I took over, John expounded his 'diminishing tenure' rule to me by drawing a little graph of duration of successive *Nature* editorships. Norman Lockyer, the first, served for a remarkable 50 years, but the stints of his successors — Richard Gregory, joint editors Jack Brimble and Arthur Gale, and John himself — became steadily shorter. In his impish way, John, who had been editor for seven years, predicted I'd last three-and-a-half.

Fortunately I managed rather longer, but when John, by then director of the Nuffield Foundation, got wind of my interest in moving on, he invited me to lunch and revealed that he very much wanted to get back into the editor's chair. Out came the imp in him again: 'Why don't we swap jobs?'

He returned in 1980; at that time, many doubted his wisdom in going back. He proved us wrong over the next 15 years and spectacularly disproved the 'diminishing tenure' rule.

David Davies Cross Keys House, Fovant, Salisbury SP3 5JH, UK

The Nature *John Maddox special is at http://tinyurl.com/dm6p7s*

Water: conflicts set to arise within as well as between states

SIR — In her Essay 'Do nations go to war over water?' (*Nature* **458,** 282–283; 2009), Wendy Barnaby quotes from my 1995 speech in Stockholm, in which I said "The wars of this century have been on oil, and the wars of the next century will be on water ... unless we change the way we manage water". The opening part was picked up by the media as a sound bite that was nevertheless valuable in pushing water issues up towards the top of the agenda, although the caveat, the operative part, was largely overlooked.

However, I do not consider that to be alarmist. I know all the arguments that have been made by others about international wars being unlikely for water, and they are probably right. But civil strife between competing groups within countries over water rights are very serious. Many of the wars of the past 20 years, on issues other than water, have been between groups within one sovereign state. That did not make them any less murderous.

Furthermore, the century is just starting and we have not seen the full range of expected environmental, demographic and political challenges unfold. Water in this century will become a major source of strife between groups within countries. Drought has driven many tribes in Africa into terrain that they are not normally expected to occupy. When coupled with other factors such as ethnic or religious divides, this becomes a dangerous mix.

Water may also become a *casus belli* between states, if the downstream nation is considerably stronger militarily than the one upstream, and the latter tries to block or reduce the flow of water. Whether it is acted on or not depends on many other issues, including the nature of the relationships between the countries concerned.

Solutions will require actions on many fronts, including in many other sectors with which water interacts economically and environmentally. But much also remains to be done to improve our resource management in the water sector broadly defined: water for food, industry, energy, domestic and municipal use, and for the environment.

The answer to the clarion call of 1995 to avoid 'water wars' is to manage our water resources better, learning from past experience, generalizing best practices and facing up to the mounting challenges that are coming our way, not to dismiss the issue as a myth.

Ismail Serageldin Library of Alexandria, Shatby 21526, Alexandria, Egypt
e-mail: is@bibalex.org

Water: resistance on the route towards a fair share for all

SIR — Wendy Barnaby's Essay 'Do nations go to war over water?' (*Nature* **458,** 282–283; 2009) is a welcome counter to mainstream media hype about conflicts over water. But all is not quiet on the waterfront, and the need to establish fair water-sharing is growing increasingly urgent.

For example, southern Iraqi farmers downstream of dams located on the Tigris River in Iraq, Syria and Turkey are being forced into urban centres as the reduced river flows become overwhelmed by sea water. Palestinian farmers eke out a living dependent on highly variable and scarce rainfall, next door to the industrial farms of Israeli settlers whose irrigation water is state-subsidized. The flood-and-drought cycles of the Ganges inundate farmers in downstream Bangladesh.

Attempts to reconcile the mockery that this fluid resource makes of political borders are well under way. The movement to establish fair water-sharing principles is gaining momentum among legal bodies and non-governmental organizations. Although the UK government is resisting calls to ratify the 1997 United Nations Watercourses Convention, demographic and anticipated climate-change pressures dilute its excuses.

Water conflicts (not wars) are a clear and present danger for millions. They deserve our full collective scientific, financial and diplomatic attention.

Mark Zeitoun School of International Development, University of East Anglia, Norwich NR4 7TJ, UK
e-mail: m.zeitoun@uea.ac.uk

ESSAY

Is free will an illusion?

Scientists and philosophers are using new discoveries in neuroscience to question the idea of free will. They are misguided, says **Martin Heisenberg**. Examining animal behaviour shows how our actions can be free.

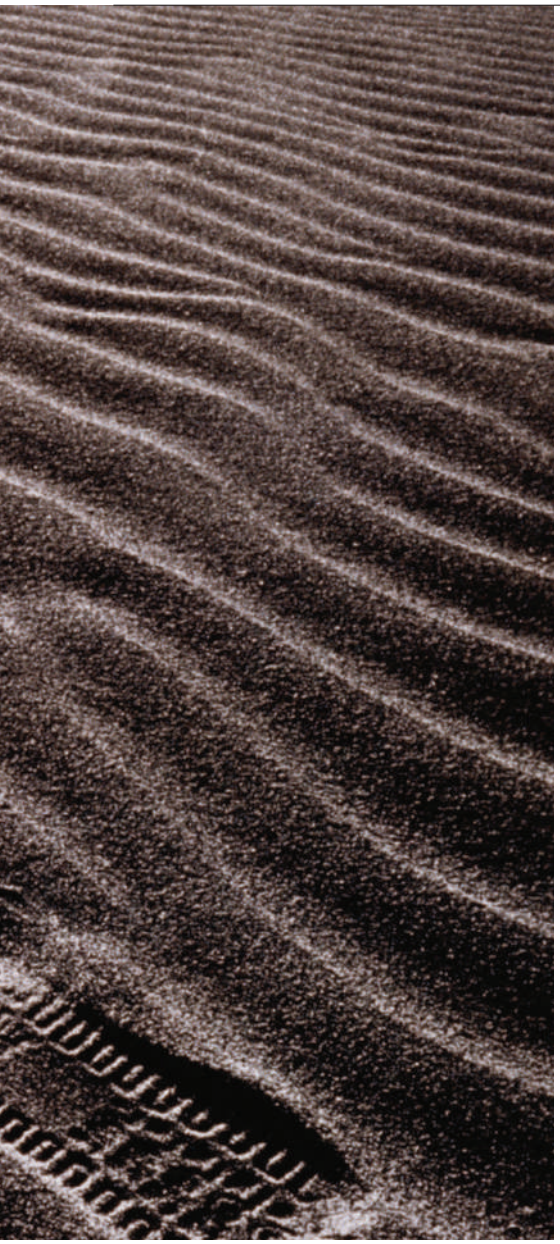
Our influence on the future is something we take for granted as much as breathing. We accept that what will be is not yet determined, and that we can steer the course of events in one direction or another. This idea of freedom, and the sense of responsibility it bestows, seems essential to day-to-day existence.

Yet it is under attack as never before. Some scientists and philosophers argue that recent findings in neuroscience — such as data published last year suggesting that our brain makes decisions up to seven seconds before we become aware of them — along with the philosophical principle that any action must be dependent on preceding causes, imply that our behaviour is never self-generated and that freedom is an illusion[1–3].

This debate has focused on humans and 'conscious free will'. Yet when it comes to understanding how we initiate behaviour, we can learn a lot by looking at animals. Although we do not credit animals with anything like the consciousness in humans, researchers have found that animal behaviour is not as involuntary as it may appear. The idea that animals act only in response to external stimuli has long been abandoned, and it is well established that they initiate behaviour on the basis of their internal states, as we do.

Before going into behaviour, I would like to take a step back and look at the nature of freedom and determinism at a more fundamental level. Almost 100 years ago, quantum physics eliminated a major obstacle to our understanding of this issue when it disposed of the idea of a Universe determined in every detail from the outset. It uncovered an inherent unpredictability in nature, in that we can never know precisely at a given moment all properties of a particle — such as both its position and its momentum.

How is this reflected at the level of everyday experience? At the scale of planets, quantum effects give way to the deterministic laws of classical mechanics. At an intermediate scale, however, they are occasionally amplified to become observable, for example when we measure radioactive decay. In general, life is an interplay between the deterministic and the random. There is plenty of evidence of chance at work in the brain: take the random



FOOTPRINTS OF BIRD AND PERSON BY PAUL WINDSOR/SPECIAL PHOTOGRAPHERS ARCHIVE/BRIDGEMAN; AKG-IMAGES

opening and closing of ion channels in the neuronal membrane, or the miniature potentials of randomly discharging synaptic vesicles. Behaviour that is triggered by random events in the brain can be said to be truly 'active' — in other words, it has the quality of a beginning.

Evidence of randomly generated action — action that is distinct from reaction because it does not depend upon external stimuli — can be found in unicellular organisms. Take the way the bacterium *Escherichia coli* moves. It has a flagellum that can rotate around its longitudinal axis in either direction: one way drives the bacterium forward, the other causes it to tumble at random so that it ends up facing in a new direction ready for the next phase of forward motion. This 'random walk' can be modulated by sensory receptors, enabling the bacterium to find food and the right temperature.

What this tells us is that behavioural output can be independent of sensory input. This is in line with the fact that in the early development of individual organisms the motor system slightly precedes the sensory system. The same may have been true in evolution, as merely being dispersed in space should have been advantageous and should have favoured mobility.

What of more complex behaviour? With the emergence of multicellularity, individual cells lost their behavioural autonomy and organisms had to reinvent locomotion. Behaviours in complex organisms typically come in modules: the grasp reflex of the newborn, the syllables of birdsong, the rhythmic motion of the legs during walking. Some modules, such as the heartbeat, last from embryonic development until death; others, such as the snapping of a crocodile's jaw, last just fractions of a second. Some can take place in parallel, like walking and singing; others are mutually exclusive, such as sleeping and playing the piano. Some necessarily follow one another, like flight and landing. From beginning to end, the lives of animals and humans are an ongoing interweaving of these behavioural modules.

Philosopher Immanuel Kant defined free will as moral, not selfish.

As with a bacterium's locomotion, the activation of behavioural modules is based on the interplay between chance and lawfulness in the brain. Insufficiently equipped, insufficiently informed and short of time, animals have to find a module that is adaptive. Their brains, in a kind of random walk, continuously pre-activate, discard and reconfigure their options, and evaluate their possible short-term and long-term consequences.

The physiology of how this happens has been little investigated. But there is plenty of evidence that an animal's behaviour cannot be reduced to responses. For example, my lab has demonstrated that fruit flies, in situations they have never encountered, can modify their expectations about the consequences of their actions. They can solve problems that no individual fly in the evolutionary history of the species has solved before. Our experiments show that they actively initiate behaviour[4]. Like humans who can paint with their toes, we have found that flies can be made to use several different motor outputs to escape a life-threatening danger or to visually stabilize their orientation in space[5].

"There is plenty of evidence that an animal's behaviour cannot be reduced to responses."

Does this tell us anything about freedom in human behaviour? Before I answer that, let's establish what I mean by freedom. One acknowledged definition comes from Immanuel Kant, who resolved that a person acts freely if he does of his own accord what must be done. Thus, my actions are not free if they are determined by something or someone else. As stated above, self-initiated action is not in conflict with physics and can be demonstrated in animals. So, humans can be considered free in their behaviour, in as much as their behaviour is self-initiated and adaptive.

Some define freedom as the ability to consciously decide how to act. I maintain that we need not be conscious of our decision-making to be free. What matters is that our actions are self-generated. Conscious awareness may help improve our behaviour, but it does not necessarily do so and is not essential. Why should an action become free from one moment to the next simply because we reflect upon it?

Kant's famous 'Third Antinomy' in his *Critique of Pure Reason* (1781) sees us on the one hand determined by natural law and on the other free because of our capacity to obey moral law. He would have been delighted to see this dilemma solved by quantum physics and behavioural biology. ■

Martin Heisenberg is professor emeritus in the department of biology at the University of Würzburg, Germany.
e-mail: heisenberg@biozentrum.uni-wuerzburg.de

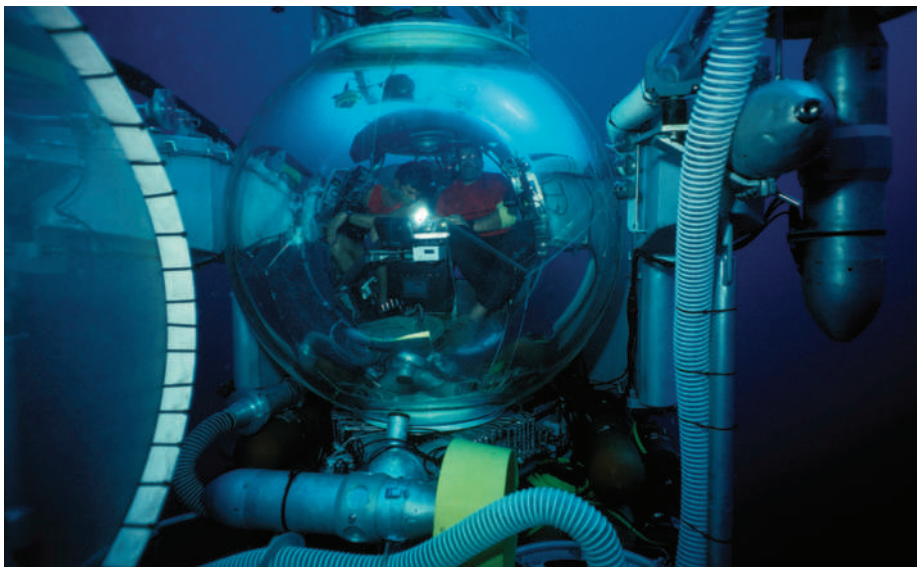
1. Soon, C. S. *Nature Neurosci.* **11,** 543–545 (2008).
2. Libet, B. *Behav. Brain Sci.* **8,** 529–566 (1985).
3. Wegner, D. M. *The Illusion of Conscious Will* (MIT Press, 2002).
4. Heisenberg, M. *Naturwissenschaften* **70,** 70–78 (1983).
5. Heisenberg, M. & Wolf, R. *Vision in* Drosophila in *Studies of Brain Function Vol. XII*, (ed. Braitenberg, V.) (Springer, 1984).

BOOKS & ARTS

The otherness of the oceans

As scientists discover more about the genomes of marine microorganisms, new views of their physiology and ecosystem networks are opening up, explain **Alexandra Z. Worden** and **Darcy McRose**.

Alien Ocean: Anthropological Voyages in Microbial Seas
by Stefan Helmreich
University of California Press: 2009.
464 pp. $60.00, £42.95



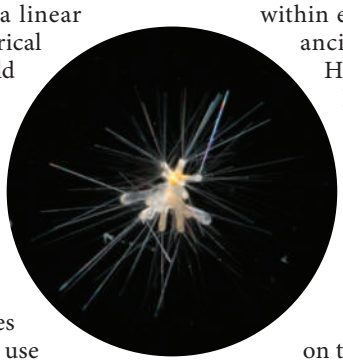
S. HADDOCK/MBARI; M. CONLIN/V&W/IMAGEQUESTMARINE.COM

Deep-sea exploration may look otherworldly, but the ocean's tiny inhabitants — such as the foraminifer *Hastigerinella digitata* (below left) — are no more alien than the microbes in our guts.

In 1913, the French essayist Marcel Proust mused that "in certain climes whole tracts of air or ocean are illuminated or scented by myriads of protozoa which we cannot see". Proust's reflections perhaps resulted from his hypochondriac perception that nature and his personal well-being were in conflict. A century later, anthropologist Stefan Helmreich examines in his new book how modern microbial oceanographers experience oceans and the diverse microbial communities that dwell within them. He proposes that microbial ocean life is 'alien' to wider society and that this 'otherness' shapes our stewardship of nature — or lack thereof.

Alien Ocean focuses on current research in marine microbiology and the social and political contexts in which it takes place, with a particular focus on metagenomics, the study of genetic material from uncultured microorganisms. Helmreich offers vignettes from his voyage through specialized research labs, conferences and oceanographic expeditions. He discusses marine biotechnology, including 'bioprospecting' for commercially valuable compounds, and social quandaries surrounding intellectual-property rights.

The book is perhaps best read as a collection of essays rather than a linear progression. It is not a historical account of the burgeoning field of microbial oceanography; for instance, the discovery of SAR11, the most abundant group of heterotrophic bacteria in the ocean surface, slips by in a single sentence. This discovery, by Stephen Giovannoni of Oregon State University and his colleagues in 1990, helped usher in the use of the polymerase chain reaction as a tool for exploring marine microbial diversity. It engendered the idea that we could, and should, pursue genetic information even in the absence of lab cultures of microbes — a central goal of the metagenomic work discussed at length in the book.



Helmreich literally dives into his research, working with remotely operated vehicles that access the deep sea as well as participating in lab work. He offers a unique glimpse into the lives of environmental scientists, whether through his underwater trips in the tiny bathysphere *Alvin* or his adventures sploshing through salt marshes with legendaries such as biologist Lynn Margulis, a proponent of the Gaia concept of linked Earth systems, who also championed the idea that mitochondria and chloroplasts within eukaryotic cells arose from ancient endosymbiotic events. Helmreich's immersion allows him to capture sentiments that are absent from publications or formal interviews. He relays scientists' fears that the cumulative damage of human interaction with the oceans will result in our own demise.

At times, Helmreich's reliance on the perspectives of his interviewees precludes achieving a balanced point of view. For example, the work of controversial biologist Craig Venter, and seemingly his personality, is attacked in a chapter formulated from interviews with Venter's critics. This is out of character with the rest of the book — the author rarely provides critical evaluation of other scientists' research, and Venter himself does not seem to have been interviewed. Helmreich also states which US labs supposedly lead the field of marine microbiology — a surprising statement, given that he does not present a comprehensive or objective evaluation of the field.

The strength of *Alien Ocean* is its innovative analysis of the ways in which oceans are alien to humans. Helmreich likens marine microbial research to voyages in outer space, but this is not always the perception among the researchers interviewed. Many of them consider the ocean and its microbial flora and fauna to be no more alien than the enormous microbial community housed within the human gut. Helmreich points out that most of society perceives things differently. For some, "the alien ocean is a medium … dense in its darkness, crushing in its pressure, suffocating in its substance", even though it hosts creatures "whose very otherness is crucial to human life support".

He drives home the point that environmental destruction by humankind is tied to this perception; the sense of oceans as being separate from our everyday lives colours our ability and willingness to defend the natural world. Thus, he warns that, "alienated from the ocean, humanity may be destined to damage it". If he can bring this conundrum into the general consciousness, it will have profound effects on how society and policy-makers interact with the oceans.

Ever present in the book is the promise microbial oceanographers place in environmental genomic sequencing. Researchers hope that this technique will lead to an "understanding [of] the genetic control of the physiology of the sea". One scientist interviewed puts forth the idea that one-third to one-half of microbial genomes might represent 'ecology genes' — genes that could explain the dynamics and interactions of organisms.

Scientists also discuss in the book how the ocean may operate as a network of genes, an increasingly popular perspective. The gene-network concept is a helpful framework for considering interactions over long timescales. But if taken too literally, its value is less clear, especially over the shorter time frames of anthropogenic perturbations. It is a microbe's entire gene complement that shapes its overall physiology, response capabilities and interactions with other life forms — not just a single gene or gene set. Hence, simply probing the oceans for genes will not necessarily provide the organism- and genome-specific context probably needed to understand microbial dynamics.

Helmreich mentions the problem of deciphering meaning from the vast amounts of information being produced at increasingly rapid rates. This disconnection between information, inference and true function really is the 'elephant in the laboratory'. How do we move beyond gene sequences to understand cell physiology, functional roles, rates, activities, trophic linkages and global biogeochemical cycles?

As much as *Alien Ocean* captures the excitement and crucial nature of oceanographic research, the field still faces the grand challenge of advancing from sequence information to a functional understanding of the system. Although proper experimental design and a statistically appropriate depth of sequencing have yet to be achieved in marine metagenomic studies, there are more fundamental issues at hand. Many 'genes' sequenced in environmental metagenomic studies, or indeed in complete genome sequences of marine bacteria, archaea and unicellular eukaryotes, are still of unknown function. Stories are woven from those genes we can name, but other genes that might render insights into how a microbe wrangles with its environment remain untouched.

Gaps in our knowledge have led to raging disputes among microbiologists about whether common measures of microbial biodiversity reflect functional divergence. The state of the field is reminiscent of periods in medical research when inferences were made about the existence and roles of particular genes and molecules, such as the tumour suppressor p53, before concrete data were available. Reactions to such inferences propelled the field forwards. In the case of microbial oceanography, a tangible forward step would be to elucidate gene function and links to physiology. This would go a long way towards moving from sequence space to ecosystem-level understanding. Perhaps *Alien Ocean* will inspire the next generation to fulfil the promise of environmental genomic sequencing. ■

Alexandra Z. Worden and **Darcy McRose** are at the Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA.
e-mail: azworden@mbari.org

See Editorial, page 140.

Ecology lost and found

Paradise Found: Nature in America at the Time of Discovery
by Steve Nicholls
University of Chicago Press: 2009.
536 pp. $30

Albert Bierstadt's 1864 work portrays the idyll of Yosemite before European settlers reached the American West.

MUSEUM OF FINE ARTS, BOSTON, 2009/BRIDGEMAN ART LIBRARY

"We don't need history," I recently heard a conservation scientist tell a group of students. He was being provocative, targeting those ecologists who treat the past as a baseline to which we should return. The world has changed too much and is changing too fast, he argued, for history to serve as a useful measure for restoring nature. The questions that animate conservation today do not ponder what we have lost or how we can get it back. The past is another world, and that world is gone. The questions now are: what kind of world do we want? And how can we create it?

In this context, Steve Nicholls's *Paradise Found* seems quaintly historical. The book is a cornucopia overflowing with the abundance of nature long gone. In this history, no species simply existed in the past. In early North America, Nicholls writes, "the fertile coasts teemed with fish and marine mammals … prairies were a carpet of wildflowers" and the mountains were "clothed in forests".

This is history written as if the past were a spectacular nature documentary. This comes as little surprise when you learn that the author has been a producer of nature shows for television for the past 25 years, as he frequently reminds readers. The book even calls to mind historical re-enactments, as Nicholls asks his audience to imagine themselves with the eleventh-century Norsemen settlers in fabled Vinland; fishing for cod with the fifteenth-century explorer John Cabot; or sitting under a tulip poplar as a flock of passenger pigeons burdens the boughs overhead.

Nicholls laments species that have been driven to extinction, but his real concern is a decline in the abundance of animals. By this he seems to mean wildlife spectacles that would be suitable for television. But Nicholls engages an important historical and ecological argument here, too. "An accurate picture of the past is important," he writes, as "a baseline to judge how effective conservation measures are."

A debate is raging among historians and ecologists regarding 'shifting baselines', a concept developed by marine biologist Daniel Pauly to describe how people often assess environmental decline only in the context of their own lifetimes. The sentiment is familiar: "When I was a kid, fish were everywhere — and this big! Now they're much harder to catch, and smaller." Each generation begins with a diminished baseline. This insight has led to massive efforts to find an accurate baseline for the natural world — what Nicholls calls 'paradise found' — by enlisting historians to scour records such as letters, diaries and ships' logs, which were also used by Nicholls to construct his narrative.

Although Nicholls acknowledges that early Native Americans often had a major impact on species and ecosystems, such as their hunting of bison in the Western grasslands, he portrays North America before its discovery by Europeans as Eden before the Fall. This is a common move in such accounts. But Nicholls seems to have

no critical distance from his own paradisiacal tropes, nor any apparent awareness that these ideas also have a history that matters.

If paradise lies in the past, it logically follows that it is lost in the present. Similarly, what is missing in the present constituted paradise in the past. This is history as elegy, and makes Nicholls's stories about catastrophic crashes in wildlife populations sound like the same "inevitable trajectory" of decline. At one point, he even apologizes for the "all too familiar pattern".

The problem is that such stories are not all the same. Some species are so successful today that they are an ecological nuisance — for example, mute swans, zebra mussels and white-tailed deer. Population size is not everything: it depends on habitat. Humans take up a lot of habitat, but we have also created new habitats, and many small populations can survive just fine. To his credit, Nicholls does not hide these complications, but he doesn't make much of them either. This makes *Paradise Found* the kind of history that undermines itself on close reading: so much complexity spills out of this bounteous tome that the narrative cannot hold it.

And that points to a much bigger problem. There is no new historical narrative to replace the simplistic story of shifting baselines and paradise lost. As a result, many ecologists are simply abandoning history. This is not good: ecology is a historical science, and history is not just data for constructing a baseline for ecological models. It unpacks everything that goes into making the baselines and models themselves — ideas, scientific theories, social practices, industries, economies, ecological conditions and species that together shaped the environment at any given time in the past. Historical narratives also frame how we think about moving forward. So they must adjust to new information, open up new inquiries, force us to rethink data and question conventional wisdom.

In many places, we have only fragments of the abundant ecosystems that once existed, and only fragments of their history. The point is not to assemble those fragments as gospel, showing the way to a past to which we might return. The point is to put this history in conversation with ecological possibilities for the present and in the future. The devil, as they say, is in the details. And we might find some useful history there too, if we could just stop searching for paradise. ■

Jon Christensen is associate director of the Spatial History Project in the Bill Lane Center for the American West at Stanford University, California 94305-4225, USA.
e-mail: jonchristensen@stanford.edu

The dangers of denying HIV

Denying AIDS: Conspiracy Theories, Pseudoscience, and Human Tragedy
by Seth Kalichman
Springer: 2009. 205 pp. $25

Inadequate health policies in South Africa have reportedly led to some 330,000 unnecessary AIDS deaths and a spike in infant mortality, according to estimates by South African and US researchers. This carnage exceeds the death toll in Darfur, yet it has received far less attention. Seth Kalichman, a US clinical psychologist, shows in *Denying AIDS* how words can kill. His marvellous book should be read alongside Nicoli Nattrass's *Mortal Combat*, covering similar ground but from the perspective of a South African.

The tragic events in South Africa have been exacerbated by AIDS 'denialists' who, Kalichman alleges, assert that HIV is harmless and that antiretroviral drugs are toxic. The author discusses the psychology of denialism, which he says is "the outright rejection of science and medicine". He recounts the history of an HIV-infected US woman whose daughter died from an AIDS-related disease, and who recently died herself, to demonstrate the downward path from "ordinary psychological denial to malignant denial to denialism". Kalichman dismisses denialists' attempts to portray themselves as intellectually honourable dissidents who question accepted wisdom. He draws clear distinctions between dissidence and denialism; the latter, he says, is merely a destructive attempt to undermine the science.

N. BOTHMA/EPA/CORBIS

South Africa's high rate of HIV infection has spurred protesters to demand action to treat it.

These attitudes are not unique to HIV. Denialism, notes Kalichman, is "partly an outgrowth of a more general anti-science and anti-medicine movement". Groups that support intelligent design, doubt global warming, claim that vaccines cause autism, argue that cigarettes are safe, believe that the terrorist attacks of 11 September 2001 were an intelligence-agency plot or deny the Holocaust all use similar tactics.

Kalichman asserts that influential groups within the AIDS denialist movement include academics, pushers of 'quack' cures and supportive journalists. He describes the academics involved as "deranged and disgruntled university professors who turn to pseudoscience as a platform to gain attention", noting that pseudoscience may include "sightings of UFOs, alien abductions, astrology, psychic predictions ... [and] outlandish claims about the cause and cure of diseases".

Kalichman describes how quacks, like some of the academics involved, misrepresent their qualifications to create an illusion of authority. One, he claims, treats AIDS with hyperthermia, massage, oxygen, music, colour, gem, aroma, hypnosis, light and magnetic fields, each word followed by "therapy". Another allegedly distributed a product in Zambia called Tetrasil, a pesticide used in swimming pools, until the Zambian government intervened. Kalichman concludes that "taking money from the poor for bogus treatments is beyond criminal" and castigates journalist supporters of the denialist viewpoint for neglecting their professional obligations to verify facts and avoid sensationalist stories. In a powerful ending, Kalichman claims that extreme right-wing politics influences the AIDS denialist movement.

Professional institutions continue to tolerate the conduct of academic denialists, despite the suffering that has resulted. The standard excuse for inaction has been freedom of expression — the First Amendment of the United States Constitution. But free speech has recognized limits, and causing death is one. In 2006, as Kalichman records, a group of concerned scientists and activists created a website, AIDSTruth (www.aidstruth.org), to provide evidence to counter the denialists' words. The international legal and human-rights communities should now investigate the deadly impact of AIDS denialism. Action might have widespread benefits: Paul Offit's tour de force, *Autism's False Prophets*, claims that pseudoscientists and quacks have used similar tactics to parasitize the suffering of desperate parents by persuading them that vaccines cause autism. As Kalichman says, denialism "will not break until the public is educated to differentiate science from pseudoscience, facts from fraud". ■

John P. Moore is professor of microbiology and immunology at the Weill Medical College of Cornell University, New York 10021, USA.
e-mail: jpm2003@med.cornell.edu

B. MCDERMOTT

Vanessa Gould was intrigued by the idea of origami as visual maths.

Q&A: Origami unfolded

In her documentary *Between the Folds*, film director **Vanessa Gould** explores the expression of mathematics through origami. She tells *Nature* how she became captivated by the art and science of transforming sheets of paper into three-dimensional geometric shapes — and exposed a hidden subculture.

Why did you make a film about origami?

I was working on Wall Street in New York, earning a living with the mathematics side of my head, but not happily. I was number crunching by day but coming home at night and painting. My degree is in physics and architecture.

Then, around five years ago, I heard about a mathematician, Tom Hull, and a computer scientist, Erik Demaine, who were using origami in their research. I was fascinated with the idea that in doing something mathematical, you could produce something beautiful to look at. A friend challenged me to make a short film about it. I had never picked up a camera before.

R. J. LANG/PHOTO BY L. GARDINER

How did you find the story?

When I visited Tom Hull in Massachusetts, I felt that I'd hit on a gold mine. He showed me an origami piece called *Five Intersecting Tetrahedra*, a beautiful, three-dimensional pointed star made with 30 pieces of paper. As I was leaving, he said 'Hey, I'd love to introduce you to a friend of mine'. His friend, a paper-maker, started talking to me about the same medium of origami but from the opposite perspective.

And that became the story — the fact that artists and scientists were all working with the same medium. Whose hands are going to hold the paper, and what are they going to turn it into?

What are your favourite shapes?

Eric Joisel folds the human form in a way that really blows audiences away. And Chris Palmer makes a spinning top out of a single square; when you pull the corners it torques the paper in such a way that it spins for 30 seconds afterwards, and that always gets a huge gasp. There's also Miyuki Kawamura's *Cosmosphere*, a huge, self-supporting sphere which is made out of many hundreds of pieces of paper.

Are there any unusual uses of origami?

We focus on a woman in Israel, Miri Golan, who has developed a mathematics curriculum which she calls Origametria. It has been extremely successful, and thousands of kids every week in Israel learn geometry through paper-folding.

What challenged you most?

It was hard to present the scientific ideas in the film without intimidating the audience. The aim was to show science in a poetic and romantic way, but with depth so it could appeal to existing scientists and maybe titillate non-scientists. Art is a metaphor for science — they are just two different lenses through which we see the Universe. ■

Interview by **Roxanne Khamsi**, news editor at Nature Medicine in New York.

See www.greenfusefilms.com for future screenings.

Art tied up

Ravelling, Unravelling
Royal Institution of Great Britain, London
Until 28 May 2009.

A chance meeting between artist Naheed Raza and mathematician Steven Bishop led to Raza's recent year-long residency in the mathematics department at University College London. Four of her resulting works, on show this month at the Royal Institution of Great Britain, examine knotted structures and the parts they play in the body and in disease, as well as in mathematical theory.

In *Nidus 1–4*, four tiny, prototype bronze casts of tangled blood vessels resemble intricate jewellery. These malformations can impede blood flow to tissues and are implicated in neurological diseases such as Alzheimer's, Creutzfeldt-Jakob and Parkinson's, and in epilepsy.

Mile of String is a rigid three-dimensional structure, made by twisting a single length of twine so that it holds a complex, coral-like form under its own tension. It evokes both Albert Einstein's concept of warped space-time and the folded and coiled structures of proteins and DNA.

For *Silk*, Raza filmed a golden orb-weaver spider. Her focus shifts between close-up shots of the spider extruding silk and hypnotic footage of its web, pulsating in the breeze. The high tensile strength of spider silk has led to its being investigated as a biomaterial that could provide a scaffold for the formation of new body tissues.

The fourth work is a digital animation produced in collaboration with Carl Fairweather. *Ravel* shows twisting, coiling ropes (pictured below) that undergo ever more complex permutations while being pulled into a vortex.

Raza says that "there is a convergent ground for fruitful dialogue about knotting as a recurring motif in science and medicine, art and culture". ■

Colin Martin is a writer based in London, UK.

N. RAZA/C. FAIRWEATHER

NEWS & VIEWS

ORIGINS OF LIFE

Systems chemistry on early Earth

Jack W. Szostak

Understanding how life emerged on Earth is one of the greatest challenges facing modern chemistry. A new way of looking at the synthesis of RNA sidesteps a thorny problem in the field.

It is well established that the evolution of life passed through an early stage in which RNA played central roles in both inheritance and catalysis[1] — roles that are currently played by DNA and protein enzymes, respectively. But where did the RNA come from?

Experiments reported by Powner *et al.*[2] (page 239 of this issue) provide fresh insight into the chemical processes that might have led to the emergence of information-coding nucleic acids on early Earth.

For 40 years, efforts to understand the prebiotic synthesis of the ribonucleotide building blocks of RNA have been based on the assumption that they must have assembled from their three molecular components: a nucleobase (which can be adenine, guanine, cytosine or uracil), a ribose sugar and phosphate. Of the many difficulties encountered by those in the field, the most frustrating has been the failure to find any way of properly joining the pyrimidine nucleobases — cytosine and uracil — to ribose[3] (Fig. 1a). The idea that a molecule as complex as RNA could have assembled spontaneously has therefore been viewed with increasing scepticism. This has led to a search for alternative, simpler genetic polymers that might have preceded RNA in the early history of life.

But Powner *et al.*[2] revive the prospects of the 'RNA first' model by exploring a pathway for pyrimidine ribonucleotide synthesis in which the sugar and nucleobase emerge from a common precursor (Fig. 1b). In this pathway, the complete ribonucleotide structure forms without using free sugar and nucleobase molecules as intermediates. This central insight, combined with a series of additional innovations, provides a remarkably efficient solution to the problem of prebiotic ribonucleotide synthesis.

The key to Powner and colleagues' approach was to overcome the deeply ingrained prejudice that carbon–oxygen chemistry (which leads to sugar formation) and carbon–nitrogen chemistry (which leads to nucleobase formation) should be kept separate for as long as possible. One does not have to look far to find the source of this prejudice. Incubation of formaldehyde — a simple carbon–oxygen compound — in alkaline solution rapidly yields a mixture of dozens of sugars[3], which subsequently react to yield an intractable tar of insoluble products. Similarly, simple carbon–nitrogen compounds, derived from cyanide and ammonia, react with each other to generate not only the standard nucleobases, but also many other compounds. It is perfectly reasonable to expect that uncontrolled mixing of these two complex processes would lead to a chemical combinatorial explosion: the synthesis of millions of different organic compounds, of which the desired biological precursor molecules would be a vanishingly small fraction. But in a remarkable example of 'systems chemistry', in which reactants from different stages of a pathway are allowed to interact, Powner *et al.*[2] show that phosphate tames the combinatorial explosion, allowing oxygenous and nitrogenous reactants to interact fruitfully.

The authors' path to RNA begins with the same starting materials used in many recent studies of prebiotic chemistry, but differs in the order in which they are combined. When the structurally simplest sugar, glycolaldehyde, reacts with the simplest derivative of cyanide and ammonia, cyanamide, a complex mixture of undesired compounds is formed. But Powner *et al.* add a third ingredient — phosphate — to the mix. In their reaction, phosphate acts as both a pH buffer and a catalyst, thereby short-circuiting the network of possible unwanted reactions and leading instead to the fast, efficient synthesis of a key intermediate known as 2-aminooxazole (Fig. 1b).

One of the goals of those developing theories of prebiotic chemistry is to identify geochemically plausible means of purifying key intermediates away from contaminants that might cause trouble in later reactions. The remarkable volatility of 2-aminooxazole suggests that it could be purified by sublimation, as it undergoes cycles of gentle warming from the sun, cooling at night (or at higher altitudes) and subsequent condensation. The compound would thus behave as a kind of organic snow, which could accumulate as a reservoir of material ready for the next step in RNA synthesis.

Phosphate continues to have several essential roles in the remaining steps of Powner and colleagues' pathway, in one case causing depletion of an undesired by-product, and in another

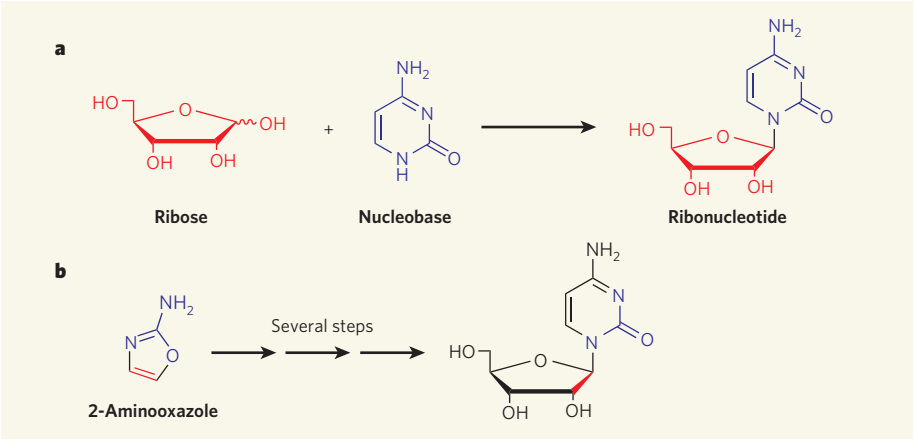


Figure 1 | Theories of prebiotic syntheses of pyrimidine ribonucleotides. The idea that RNA might have formed spontaneously on early Earth has inspired a search for feasible prebiotic syntheses of ribonucleotides, the building blocks of RNA. **a**, The traditional view is that the ribose sugar and nucleobase components of ribonucleotides formed separately, and then combined. But no plausible reactions have been found in which the two components could have joined together. **b**, Powner *et al.*[2] show that a single 2-aminooxazole intermediate could have contributed atoms to both the sugar and nucleobase portions of pyrimidine ribonucleotides, so that components did not have to form separately. For a more detailed overview of the pathways depicted here, see Figure 1 on page 239.

saving a critical intermediate from degradation. The penultimate reaction of the sequence, in which the phosphate is attached to the nucleoside, is another beautiful example of the influence of systems chemistry in this set[2] of interlinked reactions. The phosphorylation is facilitated by the presence of urea[4]; the urea comes from the phosphate-catalysed hydrolysis of a by-product from an earlier reaction in the sequence.

The authors wrap up their synthetic tour de force by using ultraviolet light to clean up the reaction mixture. They report that ultraviolet irradiation destroys side products while simultaneously converting some of the desired ribocytidine product to ribouridine (the second pyrimidine component of RNA). The development of this complex photochemistry required remarkable mechanistic insight from Powner and colleagues, who not only correctly predicted that ultraviolet irradiation would destroy the majority of the by-products, but also that the desired ribonucleotides would withstand such treatment.

The authors' careful study[2] of every potentially relevant reaction and side reaction in their sequence is a model of how to develop the fundamental chemical understanding required for a reasoned approach to prebiotic chemistry. By working out a sequence of efficient reactions, they have set the stage for a more fruitful investigation of geochemical scenarios compatible with the origin of life.

Of course, much remains to be done. We must now try to determine how the various starting materials could have accumulated in a relatively pure and concentrated form in local environments on early Earth. Furthermore, although Powner and colleagues' synthetic sequence yields the pyrimidine ribonucleotides, it cannot explain how purine ribonucleotides (which incorporate guanine and adenine) might have formed. But it is precisely because this work opens up so many new directions for research that it will stand for years as one of the great advances in prebiotic chemistry. ■

Jack W. Szostak is in the Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
e-mail: szostak@molbio.mgh.harvard.edu

1. Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 23–56 (Cold Spring Harbor Laboratory Press, 2006).
2. Powner, M. W., Gerland, B. & Sutherland, J. D. *Nature* **459,** 239–242 (2009).
3. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123 (2004).
4. Lohrmann, R. & Orgel, L. E. *Science* **171,** 490–494 (1971).

MOLECULAR MICROBIOLOGY

A key event in survival

Dave Barry and Richard McCulloch

The parasitic microorganism *Trypanosoma brucei* evades recognition by its host's immune system by repeatedly changing its surface coat. The switch in coat follows a risky route, though: DNA break and repair.

Like many other single-celled pathogens, the protozoan *Trypanosoma brucei*, which causes African sleeping sickness in humans, undergoes antigenic variation — that is, it periodically switches its variant surface glycoprotein (VSG), the molecule targeted by host antibodies. But how switching is triggered has remained largely elusive. On page 278 of this issue, Boothroyd *et al.*[1] show that a DNA double-strand break (DSB) upstream of the *T. brucei VSG* gene is the likely primary event in this process. Their results add to the few, albeit crucial, cases in which DSBs trigger developmental processes: these include mating-type switching in yeast, rearrangements of immune-system genes in humans and meiotic cell division to produce sex germ cells[2].

Antigenic switching can occur through several genetic strategies, the most common being the differential activation of an archive of silent genes and pseudogenes. Although only one gene is transcribed, from a specialized expression site, switching occurs when silent genes, or their fragments, are duplicated in the expression site by a gene-conversion process, replacing all or part of the expressed gene. In some pathogens, the expressed gene can be constructed as a mosaic from several archival pseudogenes; such a combinatorial strategy expands the scale of variation enormously, with, for example, five pseudogenes giving rise to hundreds of combinations[3].

Trypanosoma brucei has evolved an even more staggeringly complex system. It, too, transcribes a single *VSG* gene, but the sources of sequences that contribute to switching are large and diverse. It has several inactive expression sites, and its archive contains up to 200 *VSG* genes that lie at the ends (telomeres) of a set of mini-chromosomes, as well as a further 1,600 silent genes — of which two-thirds are pseudogenes — on the main chromosomes[4]. The potential for mosaic variation therefore seems beyond estimation. Intact archival genes are duplicated starting from an upstream set of repeat sequences each 70 base pairs (bp) long[5], all the way to sequences at the downstream end of the coding sequence, or, in the case of silent telomeric genes, perhaps to the nearby end of the chromosome. As gene conversion in other organisms is initiated by a DSB in the conversion site, such a break has been proposed also to occur in the *T. brucei VSG*

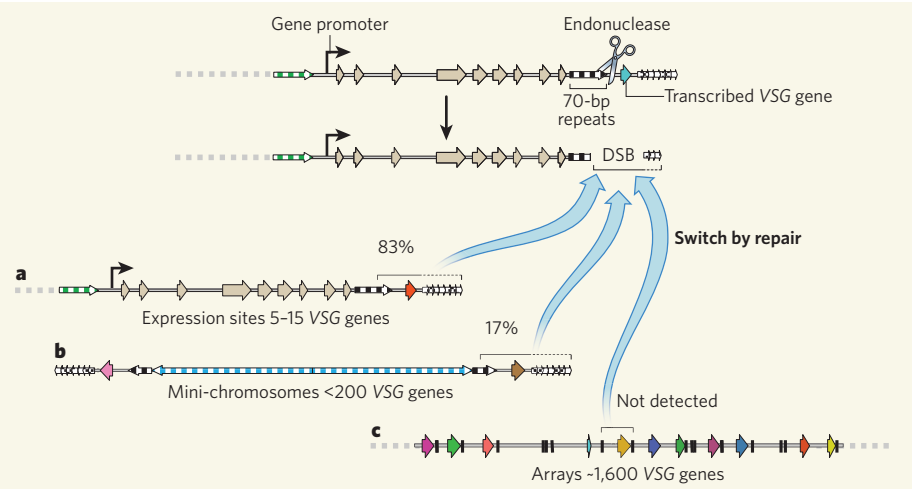


Figure 1 | Antigenic switching and sources. Boothroyd *et al.*[1] used an endonuclease enzyme to induce a DNA double-strand break (DSB) adjacent to the 70-bp-repeat region of the active *VSG* gene in *Trypanosoma brucei*. Consequently, the region from the DSB site to the end of the *VSG* gene was deleted. The protozoan filled this gap by a repair process, using silent *VSG* loci on other chromosomes as template. Locations of donor sequences included: (**a**) expression sites (of which there are 5–15 per strain) at the telomeres of the main chromosomes; (**b**) telomeres of some 100 mini-chromosomes found in the *T. brucei* genome; and (**c**) tandemly arrayed *VSG* genes in the main chromosomes. The copied regions stretched from the 70-bp-repeat regions to the telomere, or, for intact genes, to the end of the *VSG*. The frequencies of conversions the authors detected (shown as percentages) differ from those observed during infections with natural strains of *T. brucei*, in which mini-chromosomes dominate as donors. Brackets denote the duplicated region, with dashed sections indicating uncertainty over where the duplication ends. Broad arrows indicate genes; narrow arrows, repetitive DNA sequences (70-bp repeats are shown in black and white). Coloured arrows are different *VSG* genes; grey arrows, genes other than *VSG*.

expression site, in its long set of 70-bp repeats.

Boothroyd *et al.*[1] tested this hypothesis by creating a unique target site for the yeast endonuclease enzyme I-SceI in the expression site of the *T. brucei VSG* gene, adjacent to the 70-bp repeats (Fig. 1). Inducing this enzyme to become active, which caused DSBs in some 1% of trypanosomes, led to a dramatic increase in antigen switching. The switches involve typical conversions by telomeric archival *VSG* genes, stretching from the 70-bp repeats to, possibly, the chromosome end. To ascertain that the conversion was not merely repair in response to the artificial introduction of a break, the authors demonstrate that DSBs also occur naturally in the repeat regions of the transcribed gene, but seldom in another, inactive, expression site.

These findings suggest a model for switching in which natural breaks occur in the active expression site, precipitating conversion repair from another locus that contains a distinct but inactive *VSG* gene. Any model raises questions, and in this case two must be addressed.

One question is how the breaks occur. There are several possible mechanisms. First, they might be caused by an endonuclease, but such an enzyme would have to be strictly regulated to prevent lethally extensive cleaving of the many available 70-bp repeats in the genome. Second, a DNA-modification repair process might occur, similar to that mediated by the AID enzyme in human immunoglobulin-class switching[6]. But such a process seems too complex for the requirements of the *VSG* gene.

A third possibility is transcription-associated breakage. Indeed, the actively transcribed *VSG* expression site displays single-strand DNA sequences[7], which could lead to DSBs during experimental DNA isolation; such artefactual breaks, however, are unlikely in the elegant technique Boothroyd *et al.* used. Alternatively, transcription might induce instability among the 70-bp repeats, triggering repair processes that cause DNA breakage[8].

Finally, the structurally unstable repeats could also stall DNA replication, creating DSB-like free DNA ends that prompt repair through recombinational mechanisms[9]. The simplicity of this last mechanism is attractive, and could explain the abundance of 70-bp repeats in the expression site — to favour such accidents. Many bacterial phase-variation systems, which switch between alternative virulence states, do so through strategically located accident black-spots of unstable DNA-sequence tracts[10].

Another, broader question arising from the model based on Boothroyd and colleagues' observations will probably be answered only in the longer term: can initiation of *VSG* switching by DSB formation explain the complexity and hierarchy of antigenic variation? Hierarchical gene expression — a key element of antigenic variation — arises from variations in the probability that different donor *VSG* genes are activated, and this is thought to relate to locus type, flanking sequences and, in the case of mosaic genes, the presence of related silent genes. Do these various types of *VSG* switching all involve DSBs?

Of necessity, the authors have studied a laboratory-adapted trypanosome strain, in which antigenic variation is impaired both quantitatively and qualitatively — recombination events are less frequent and less specific with regard to 70-bp repeats. This compromise might explain an anomaly in their observations: virtually all sequence donors to the switch were inactive expression sites, rather than mini-chromosomal sequences, which are favoured during natural infections (Fig. 1). The presence of DSBs in the low-switcher laboratory strain implies that either the high-switcher natural strains incur many more such breaks, or an essential downstream step, possibly one involved in repair, can become defective during laboratory adaptation, leading to less frequent switching.

It is now crucial to determine whether and how DSBs yield hierarchy in natural strains. Boothroyd and colleagues' findings — which provide a testable model for phenotype-switching systems in other organisms — should also prompt researchers to investigate the molecular players involved in switching. ■

Dave Barry and Richard McCulloch are at the Wellcome Centre for Molecular Parasitology, University of Glasgow, Glasgow G12 8TA, UK.
e-mails: j.d.barry@bio.gla.ac.uk; rmc9z@udcf.gla.ac.uk

1. Boothroyd, C. E. *et al. Nature* **459,** 278–281 (2009).
2. Shrivastav, M. *et al. Cell Res.* **18,** 134–147 (2008).
3. Futse, J. E. *et al. Mol. Microbiol.* **57,** 212–221 (2005).
4. Berriman, M. *et al. Science* **309,** 416–422 (2005).
5. Aline, R. Jr *et al. Nucl. Acids Res.* **13,** 3161–3177 (1985).
6. Stavnezer, J. *et al. Annu. Rev. Immunol.* **26,** 261–292 (2008).
7. Greaves, D. R. & Borst, P. *J. Mol. Biol.* **197,** 471–483 (1987).
8. Lin, Y., Hubert, L. Jr & Wilson, J. H. *Mol. Carcinog.* **48,** 350–361 (2009).
9. Labib, K. & Hodgson, B. *EMBO Rep.* **8,** 346–353 (2007).
10. Moxon, R. *et al. Annu. Rev. Genet.* **40,** 307–333 (2006).

ASTROPHYSICS

Cosmic crystals caught in the act

Aigen Li

The outburst of a Sun-like star offers a rare opportunity to witness the making of silicate crystals in the star's planet-forming disk, providing key information about the formation of comets and the Solar System.

We live in a dusty Universe. Dust is a ubiquitous feature of the cosmos, and impinges directly or indirectly on most fields of modern astronomy. The most common cosmic-dust species — the silicates — occurs in a wide variety of astrophysical environments, ranging from comets and protoplanetary disks (planet-forming dust disks around young stars) to the most distant galaxies known, which formed when the Universe was just a few hundred million years old. The way in which atoms in silicate grains are arranged — that is, whether they are arranged in a random manner or in an ordered lattice structure, as in their amorphous and crystalline forms, respectively — provides information about their origin, in particular about their parent regions.

The origin of crystalline silicates in comets has been a matter of debate since their first detection 20 years ago[1]. Crystalline silicates are unexpected if comets are, as is widely believed, remnants of primordial material from the cold, outer parts of the protoplanetary dust disk from which the Solar System has formed, the solar nebula[2]. Although it is recognized that comets do evolve during their storage in the far reaches of the Solar System[3], they are undoubtedly the most pristine bodies in the Solar System.

On page 224 of this issue, Ábrahám *et al.*[4] present the first convincing evidence for the formation of crystalline silicates through thermal annealing of amorphous silicates in the hot, inner disk around an eruptive star. In a complementary study, Vinković[5] (page 227) proposes a viable, novel mechanism to transport the newly formed crystalline silicates from the hot, inner regions of protoplanetary disks to their cold, outer, comet-forming regions. Together, these results[4,5] offer a solution to the long-standing puzzle of the origin of crystalline silicates in comets and protoplanetary disks, and provide insight into the formation of comets and planetary systems.

As far as comet-formation theory goes, comets formed in the cold, outer regions of the solar nebula, at distances of at least 5 astronomical units (AU) from the Sun (1 AU is the distance from Earth to the Sun). They have been stored in reservoirs as far as 30–10,000 AU from the Sun, and, having formed early in the life of the Solar System, which is about 4.5 billion years old, have remained cold ever since. In observed samples of comets, the presence of highly volatile, frozen molecules, such as carbon monoxide, and molecular nitrogen (which in comets is a rare gas species), indicates that comets formed at very low temperatures, as low as about 30 kelvin. Moreover, the remarkable similarity between such volatile ices — in particular those of water, carbon monoxide, ammonia and methane — in interstellar material and in comets strengthens the link between comets and the pristine interstellar materials of the solar nebula[3].

Recently, crystalline silicates were identified

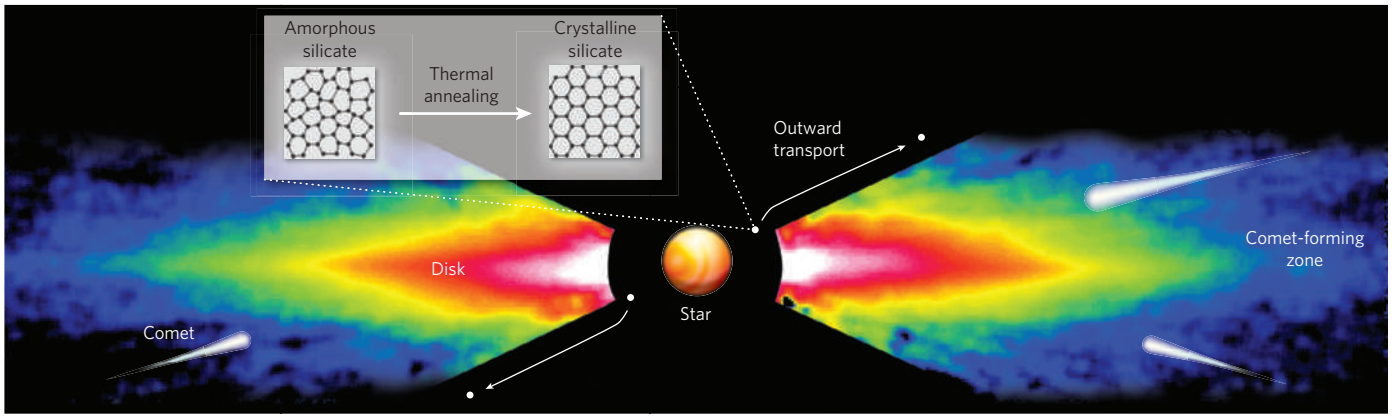


IMAGE: C. BURROWS/J. KRIST/STSCI/ESA/WFPC2 IDT TEAM/NASA

Figure 1 | Origin and transport of silicate crystals in planet-forming dust disks. According to the latest findings[4,5], crystalline silicates are produced in the hot, inner regions of a star's disk by thermal annealing of amorphous silicates, and are subsequently transported to the cold, outer comet-forming regions, where they settle towards the disk's mid-plane and are incorporated into comets. In this artist's representation, the structures of the star, comets and silicates are overlaid on a (false-colour) visible-light Hubble Space Telescope view of the star Beta Pictoris's dust disk, which is seen tilted almost edge-on from the telescope's vantage point. The disk extends to more than 1,000 AU from the star, with an inner dust-free hole of a few AU (presumably created by planets)[15]. The dark, circle-shaped region around the star is the result of light — originating from the star — being blocked by the coronagraph on Hubble.

in the dust samples collected from comet 81P/Wild 2 by the Stardust spacecraft[6]. Their presence in other comets has also been revealed by infrared (IR) spectral signatures[7]: the IR spectra display sharp emission features at several specific wavelengths that are characteristic of crystalline silicates. These distinct emission features are also seen in protoplanetary disks around young stars[8], suggesting a similar origin for crystalline silicates in comets and in these dust disks.

So where did the observed silicate crystals in comets come from? Apparently, they were not inherited from the interstellar medium, simply because interstellar silicates are predominantly amorphous[9]. They clearly did not form in cometary nuclei, which are believed to be assembled at temperatures below 30 K (ref. 2), or in the cold, outer regions of the solar nebula, where comets were accreted about 4.5 billion years ago and where materials have never experienced temperatures higher than 100 K. The crystallization of the original, amorphous silicates through thermal annealing requires temperatures of at least ~1,000 K (ref. 10), in contradiction with the scenario in which comets were formed and stored in cold environments[2].

The standard speculation has been that the volatile ices and crystalline silicates found in comets are of different origins. Whereas volatile ices may be pristine interstellar material surviving from the time the Solar System formed, crystalline silicates can originate from amorphous silicates that were transformed to crystalline form by thermal annealing in the hot, inner solar nebula, and were then transported outwards and incorporated into comets (Fig. 1).

However, it is Ábrahám *et al.*[4] who provide the first concrete observational evidence for thermal annealing of amorphous silicates. They present mid-IR spectra, in the 5.2–37-micrometre wavelength range, of the star EX Lupi, obtained at two epochs separated by an interval of about 3 years. EX Lupi is a prototypical young Sun-like eruptive star that undergoes large, repetitive outbursts. The first-epoch spectrum, obtained when EX Lupi was in a quiescent phase, displays a broad, smooth 9.7-μm emission band, a telltale signature of amorphous silicates. By contrast, the second-epoch spectrum, acquired when the star was in the middle of an outburst, exhibits several sharp peaks characteristic of crystalline silicates superimposed on the broad, 9.7-μm band of amorphous silicates. These features are similar to those observed in comet spectra.

Ábrahám *et al.*[4] interpret the observations as ongoing crystal formation: crystalline silicates are produced by thermal annealing in the surface layer of the star's inner disk (about 0.5 AU from the star) by heat from the outburst, which increases the visual brightness of the star by a factor of about 100. Alternative explanations for the observed spectral peaks, such as illumination of existing crystals residing in outer disk areas or the stirring up of crystals from the disk's mid-plane, are ruled out by modelling.

So the question that naturally arises is how these newly produced silicate crystals are carried outwards from the inner, crystal-formation zone to the cold, comet-forming zone to be incorporated into comets? Several mechanisms have been suggested, including turbulent mixing[11] of dust grains in the mid-plane of the solar nebula and the 'X-wind' model, in which dust grains are ballistically launched above the disk's mid-plane and transported

MICROBIOLOGY

Signals for change

The strategy that the protozoan parasite *Trypanosoma brucei* — which causes fatal disease in humans and cattle — uses to evade its host's immune defences is the subject of a News & Views article on page 172. Elsewhere in this issue, Dean *et al.* provide further molecular insights into the workings of this pathogen (pictured, with red blood cell). They focus on the cell-differentiation events associated with its transmission between host and vector (S. Dean *et al.* *Nature* **459,** 213–217; 2009).

In response to the metabolites citrate or *cis*-aconitate, trypanosomes differentiate from a non-dividing, stumpy form — the form thought to be taken up by the parasite's tsetse fly vector from the mammalian host's bloodstream — to a dividing form found in the fly's midgut.

The parasite becomes sensitive to the metabolite signals through exposure to low temperatures, which the fly often experiences while feeding at dusk or dawn. But the surface molecules responsible for transmitting the signals to the microbe have remained elusive.

Dean and colleagues find that trypanosomes sense citrate through the PAD family of cell-surface transporter proteins. Only the stumpy form seems to express these proteins, establishing it as the competent stage in the parasite's life cycle for transmission from its mammalian host to the fly.

EYE OF SCIENCE/SPL

The authors also find that reducing PAD expression decreases citrate-induced trypanosome differentiation. PAD proteins could therefore potentially be used as molecular markers when screening for compounds that promote transition to the non-dividing, and so less-virulent, stumpy form.

Sadaf Shadan

outwards[12]. But these models seem to have difficulty in explaining the observed levels of transport[13]. Vinković[5] proposes a transport model based on the non-radial component of the radiation-pressure force. He shows that the radiation pressure from the star, combined with that from the disk's near-IR light, could push grains outwards along the disk's surface irrespective of its curvature.

But Vinković's theory is valid for micrometre-sized dust grains, and crystalline-silicate grains that big cannot emit much light at their characteristic mid-IR wavelengths. If only micrometre-sized silicate crystals were transported to the outer disk regions, neither protoplanetary disks nor comets would exhibit the observed sharp emission features of crystalline silicates. It would be interesting to see whether other mechanisms such as turbulent mixing and the 'X-wind' model would effectively carry submicrometre grains, which are efficient mid-IR emitters, outwards and incorporate them into comets. It is also possible that some — but not all — crystalline silicates are made *in situ* in cometary comae[14]. ■

Aigen Li is in the Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211, USA.
e-mail: lia@missouri.edu

1. Campins, H. & Ryan, E. V. *Astrophys. J.* **341,** 1059–1066 (1989).
2. Crovisier, J. *Faraday Discuss.* **133,** 375–385 (2006).
3. Stern, S. A. *Nature* **424,** 639–642 (2003).
4. Ábrahám, P. *et al. Nature* **459,** 224–226 (2009).
5. Vinković, D. *Nature* **459,** 227–229 (2009).
6. Brownlee, D. *et al. Science* **314,** 1711–1716 (2006).
7. Wooden, D. H. *Space Sci. Rev.* **138,** 75–108 (2008).
8. Mann, I., Köhler, M., Kimura, H., Cechowski, A. & Minato, T. *Astron. Astrophys. Rev.* **13,** 159–228 (2006).
9. Kemper, F., Vriend, W. J. & Tielens, A. G. G. M. *Astrophys. J.* **609,** 826–837 (2004).
10. Hallenbeck, S. L., Nuth, J. A. & Daukantas, P. L. *Icarus* **131,** 198–209 (1998).
11. Bockelée-Morvan, D., Gautier, D., Hersant, F., Huré, J.-M. & Robert, F. *Astron. Astrophys.* **384,** 1107–1118 (2002).
12. Shu, F. H., Shang, H. & Lee, T. *Science* **271,** 1545–1552 (1996).
13. Ciesla, F. J. *Science* **318,** 613–615 (2007).
14. Yamamoto, T. & Chigai, T. in *Highlights of Astronomy* Vol. 13 (ed. Engvold, O.) 522–524 (Astron. Soc. Pacific, 2005).
15. Golimowski, D. A. *et al. Astron. J.* **131,** 3109–3130 (2006).

ARCHAEOLOGY

Origins of the female image

Paul Mellars

Discovery of the sexually explicit figurine of a woman, dating to 35,000 years ago, provides striking evidence of the 'symbolic explosion' that occurred in the earliest populations of *Homo sapiens* in Europe.

On page 248 of this issue[1], Nicholas Conard describes an archaeological discovery of considerable significance — arguably the world's oldest depiction of a human figure, carved in impressive detail from a solid piece of mammoth ivory, and only 60 millimetres long. The find (Fig. 1) is remarkable for several reasons.

Fragments of the figure were excavated from archaeological deposits in the Hohle Fels cave in south Germany, dated by a range of more than 30 radiocarbon measurements to at least 35,000 years in age (in terms of the newly 'calibrated' radiocarbon timescale). They were recovered in association with characteristic stone, bone and ivory tools belonging to a period, the Aurignacian, that represents the earliest settlement of Europe by fully anatomically and genetically modern human populations, and which saw the simultaneous demise of the preceding Neanderthals[2,3]. And the figure is explicitly — and blatantly — that of a woman, with an exaggeration of sexual characteristics (large, projecting breasts, a greatly enlarged and explicit vulva, and bloated belly and thighs) that by twenty-first-century standards could be seen as bordering on the pornographic. As if to emphasize the sexual characteristics, the figure's arms and legs are severely reduced in size, and the 'head' has been reduced to the form of a carefully carved ring, evidently to allow the figure to be suspended from a string or thong.

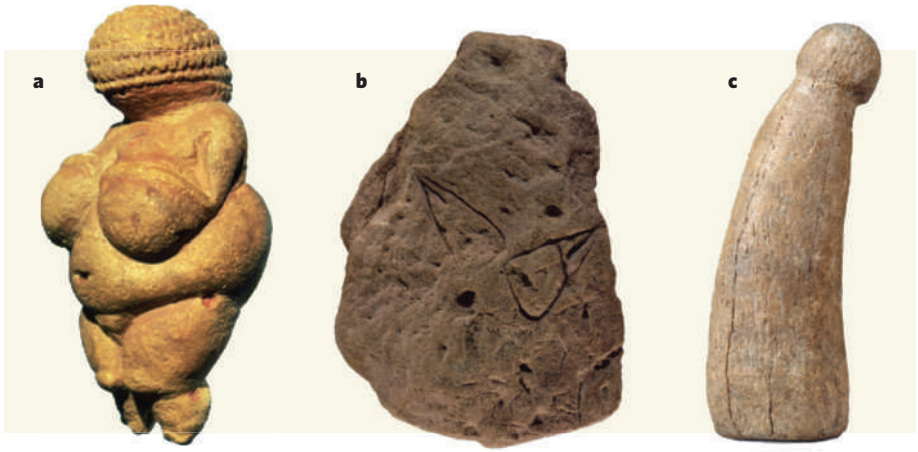
H. JENSEN/UNIV. TÜBINGEN

Figure 1 | A 35,000-year-old sex object. The newly described[1] Aurignacian figurine, 60 millimetres in height, viewed from different angles.

This find is the latest discovery in a veritable art gallery of early 'modern' human art recovered over the past 70 years from a series of cave sites located in the Schwabian region of southern Germany, only a short distance north of the Danube valley[2,4] — the route by which the earliest populations of *Homo sapiens* probably penetrated central and western Europe[3]. Four sites in this region have now produced a total of 25 small carvings, all made from mammoth ivory and depicting various forms ranging from superbly sculpted mammoths and horses, through bison and cave lions, to elegant bird-like forms, and two curious half-animal, half-human ('therioanthropic') figures[2]. The same sites have also yielded numerous small, carved ivory beads or pendants and the world's oldest unmistakable musical instruments: these take the form of perforated flutes manufactured from segments of bird wing bone and meticulously conjoined segments of mammoth ivory[4]. As a reflection of the artistic creativity of the earliest *H. sapiens* populations in Europe, this collection of south German material is currently unique.

What makes the German finds especially remarkable is their emphasis on fully in-the-round sculptures (figurines), frequently embellished with enigmatic, evidently symbolic, markings. Such markings take the form of criss-cross designs or (in the case of the newly discovered figure) repeatedly incised lines that might conceivably represent schematic depictions of skin clothing[2]. Other kinds of art forms have been known for some time from broadly contemporaneous sites in western and southern France, including — most spectacularly — the highly sophisticated drawings of horses, bison, deer, rhinos, cave lions and other animals in the Chauvet cave in southeastern France. The drawings were discovered in 1994, and dated by radiocarbon-accelerator measurements of the charcoal actually used to make the drawings to approximately 36,000–37,000 (calibrated) years ago[5]. Possibly slightly earlier in date are several paintings executed in red iron oxide on limestone slabs from the Fumane cave in northeastern Italy — including one figure that has been interpreted as an apparently quasi-human figure with animal-like horns[6]. But the cornucopia of small, carved ivory statuettes from the south German sites must be seen as the birthplace of true sculpture in the European — maybe global — artistic tradition.

The feature of the newly discovered figure that will undoubtedly command most attention is its explicitly, almost aggressively, sexual nature, focused on the sexual characteristics of the female form. As Conard[1] points out, this figure is strongly reminiscent of the later,



A. BRIDGEMAN ART LIBRARY; B, P. JUGIE/MNP, LES EYZIES, DISTR. RMN; C. RMN

Figure 2 | Sexual images in early *Homo sapiens* European art. a, A 'Venus' figurine from Willendorf, Austria, 105 millimetres in height and dated to about 28,000 years ago. Note the similarities to the older figurine from Hohle Fels, described by Conard[1] and shown in Figure 1. **b**, Female 'vulvar' symbols carved on a limestone block from the La Ferrassie rock shelter, southwest France, dating to about 35,000 years ago. **c**, A phallus, carved from the horn core of a bison, from the Blanchard rock shelter, southwest France; the carving is about 36,000 years old and is 250 millimetres long.

well-known 'Venus' figurines recovered from a range of sites stretching from the Pyrenees into southern Russia, and associated with the subsequent Gravettian toolmaking cultures. These figurines are dated to between about 29,000 and 25,000 years ago, and most of them show a similar exaggeration of the sexual characteristics and a curious downplaying of the arms, legs and heads[2,7] (Fig. 2a). The extension of this obsession with female characteristics back to at least 35,000 years ago should perhaps not come as any surprise, because explicit representations of female 'vulvar' symbols had already been recorded from a number of early Aurignacian sites in western France, all incised on blocks of limestone, and again dated back to at least 35,000–36,000 years ago[2] (Fig. 2b). Interestingly, this sexual-symbolism aspect of the art is effectively symmetrical, as the same sites have yielded equally explicit phallic representations, carved out of bone, ivory or (in one case) the horn core of a bison (Fig. 2c). The possibility that these could represent 'girls' toys' (as one first-year student once hesitantly expressed it) should perhaps not be dismissed.

Whichever way one views these representations, it is clear that the sexually symbolic dimension in European (and indeed worldwide) art has a long ancestry in the evolution of our species. To some, this has often been taken as a possible reflection of fertility beliefs, designed to ensure the continuity of life in both the human and animal realms[8]. The archaeologist and ethnographer André Leroi-Gourhan interpreted the whole of European cave art during the Upper Palaeolithic, roughly 40,000 to 15,000 years ago, in terms of a dualistic, 'structuralist' reflection of the opposition of the sexes[8]. Other workers, such as David Lewis-Williams, have seen the same symbols as possible elements in shamanistic rituals and beliefs[9].

From an evolutionary perspective, of course, the most striking feature is the sudden eruption of all these forms of artistic or other explicitly symbolic creations with the arrival of the earliest *H. sapiens* populations in Europe, and the shortly ensuing demise of the pre-existing Neanderthal populations of the continent[3]. We know that these modern populations came into Europe from Africa, where they had originated much earlier and where early forms of symbolic expression have been found as abstract, geometrical designs engraved on pieces of red iron oxide extending back to at least 75,000, and possibly 95,000, years ago[10]. But the advent of fully representational, 'figurative' art seems at present to be a European phenomenon, without any documented parallels in Africa or elsewhere earlier than about 30,000 years ago[11]. How far this 'symbolic explosion' associated with the origins and dispersal of our species reflects a major, mutation-driven reorganization in the cognitive capacities of the human brain — perhaps associated with a similar leap forward in the complexity of language — remains a fascinating and contentious issue[12,13]. ■

Paul Mellars is in the Turkana Basin Institute, Stony Brook University, USA, and the Department of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK.
e-mail: pam59@cam.ac.uk

1. Conard, N. J. *Nature* **459,** 248–252 (2009).
2. Bosinski, G. *Homo Sapiens* (Errance, 1990).
3. Mellars, P. *Nature* **439,** 931–935 (2006).
4. Conard, N. J. & Bollus, M. *J. Hum. Evol.* **44,** 331–371 (2003).
5. Clottes, J. (ed.) *La Grotte Chauvet: L'Art des Origines* (Seuil, 2001).
6. Broglio, A. & Dalmeri, G. (eds) *Pitture Paleolitiche nelle Prealpi Venete* (Museo Civico di Storia Naturale di Verona, Verona, 2005).
7. Roebroeks, W. *et al.* (eds) *Hunters of the Golden Age* (Analecta Praehistorica Leidensia, Leiden, 1999).
8. Ucko, P. J. & Rosenfeld, A. *Palaeolithic Cave Art* (Weidenfeld & Nicolson, 1967).
9. Lewis-Williams, D. *The Mind in the Cave* (Thames & Hudson, 2002).
10. Henshilwood, C. S. *et al. J. Hum. Evol.* (in the press).
11. Vogelsang, R. *Middle-Stone-Age-Fundstellen in Südwest-Namibia* (Heinrich-Barth-Inst., Cologne, 1998).
12. Mithen, S. *The Prehistory of the Mind* (Thames & Hudson, 1996).
13. Klein, R. G. *Evol. Anthropol.* **17,** 267–281 (2008).

50 YEARS AGO

A recent issue of the *Australian Museum Magazine* is devoted almost entirely to New Guinea ... The physical geography is described by D. F. McMichael and the geology by G. A. U. Stanley. J. S. Womersley discusses the vegetation of the island, while other contributors provide details about the mammals, birds, fishes and insects. Until the early 1930's it was thought that the central region of New Guinea was uninhabited and uninhabitable. Since that time it has become known that about 600,000 people live in the Australian territories alone ... This issue is also of interest for its reference to the discovery of a rare animal in Australia, the potoroo (*Potorous tridactylus*). This animal, which is related to the rat-kangaroos, is now rare in New South Wales, not having been recorded in the State since 1913. It is still common in Tasmania. The specimen obtained by the Museum was killed by a dog near Gosford, New South Wales.
From *Nature* 16 May 1959.

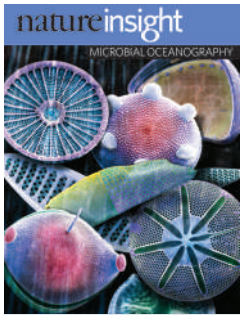
100 YEARS AGO

In the April number of *Das Blaubuch* Dr. T. Zell discusses the question whether animals take advantage of experience and become cleverer than their parents, the question being answered in the affirmative. Among numerous other instances mentioned by the author, reference may be made to the following. From early times it has been noticed that vultures have learnt to accompany armies in the field, for the sake of the prospective feast after a battle. Killer-whales accompany whaling-vessels, and gulls do the same ... Birds and quadrupeds have learnt to take no notice of railway trains, as have horses of motors, and nowadays many fewer birds immolate themselves by flying against telegraph-wires than was formerly the case ... Sheep-dogs, again, know by experience that it is only the members of their masters' flocks that it is their business to collect.
From *Nature* 13 May 1909.

50 & 100 YEARS AGO

nature insight

MICROBIAL OCEANOGRAPHY



Cover illustration
The silicon structure of the cell walls of various diatom species. (Diatom images courtesy of R. Crawford and F. Hinz. Artwork by N. Spencer.)

MICROBIAL OCEANOGRAPHY

It is now commonly accepted that the world is changing as a result of human activity. The rise in atmospheric carbon dioxide — which increases the amount of CO_2 dissolved in the ocean and reduces the pH of the water — as well as higher temperatures will probably have a detrimental effect on the oceans' ecosystems. To predict future changes, we need to understand the chemistry and biology of the marine world at present.

The smallest but arguably most important inhabitants of the ocean are the microorganisms. These organisms are at the bottom of the marine food web, they outnumber all other marine species by orders of magnitude, and are therefore central to all nutrient cycles. But their small size, the inaccessibility of their habitats, the diversity and interdependence of microbial communities, and our inability to adapt them to life in the laboratory have made them difficult to study.

Advances in large-scale genomic analyses have circumvented some of these problems and have allowed us to determine the composition of microbial communities, as well as their activity at a particular site at a given time. Robotic devices can now also incorporate time series and spatial gradients.

These efforts have provided us with interesting and surprising insights into microbial life in the ocean. In many cases, however, they have also revealed how little is known. This Insight provides a snapshot of today's research efforts in the field of microbial oceanography. It suggests that future work might not only uncover unexpected and unusual species, habitats and interactions, but also help us to understand and respond to the challenges of global change and its effect on human life.

Claudia Lupp, Senior Editor

COMMENTARY

REVIEWS

nature insight

NATURE|Vol 459|14 May 2009|doi:10.1038/nature08056

Microbial oceanography in a sea of opportunity

Chris Bowler[1,2], David M. Karl[3] & Rita R. Colwell[4]

Plankton use solar energy to drive the nutrient cycles that make the planet habitable for larger organisms. We can now explore the diversity and functions of plankton using genomics, revealing the gene repertoires associated with survival in the oceans. Such studies will help us to appreciate the sensitivity of ocean systems and of the ocean's response to climate change, improving the predictive power of climate models.

The pursuit of knowledge of the oceans has progressed in recent years thanks to the availability of new technologies and tools. Satellites are now equipped with sensors that can measure the optical properties of surface waters; profiling floats in the oceans can collect physical and chemical data from around the world; and fluorescence detectors can provide information about chlorophyll concentrations at depths beyond the reach of satellite-based sensors[1]. Remotely operated autonomous undersea vehicles have also been incorporated into oceanographic research and have been used to explore hydrothermal vents and other remote habitats. These technologies have stimulated a renewed interest in ocean exploration, and the vast quantities of physicochemical data collected have aided in the development of predictive models for some of the major ocean processes[2,3].

By contrast, the integration of cell biology and genomics into oceanographic research is much less developed, even though biological (and especially microbiological) processes are fundamental for maintaining a functional global ecosystem. Microscopic life is ubiquitous in the oceans (Box 1), and functional studies of these organisms are transforming our view of the processes and diversity of life in the world's oceans (see the other articles in this Insight). In this Commentary, we describe the crucial roles of marine microorganisms in maintaining the well-being of our planet, and we discuss how new technologies in the biological sciences can be recruited into oceanography to improve our knowledge of these processes.

Microbial diversity and evolution

Ancient microorganisms that evolved in the oceans helped to create the conditions under which more complex life developed[4]. The appearance of photosynthesis more than two billion years ago helped to shape the chemical environment that allowed the evolution of multicellular organisms and complex biological communities, including human societies. The metabolism of marine microorganisms continues to maintain major biogeochemical cycles that other organisms cannot complete, including significant production of the oxygen required for aerobic life (Box 2). For example, although terrestrial plants make up the vast majority of photosynthetic biomass on the planet, marine phytoplankton carry out almost half of the global net photosynthesis[5]. The relatively high rate of photosynthesis per unit of biomass for marine phytoplankton, compared with terrestrial plants, derives from their rapid rates of metabolism and turnover[6,7]. These facets have implications for the potential response time of microbial assemblages to climate variability and change, and for the neutralization of anthropogenic pollutants.

The oceans contain environments that resemble those that first nurtured life on Earth, such as marine sediments with marked layering of redox potentials, methane seeps from the deep subsurface, elevated heat and pressure around hydrothermal vents, and anaerobic, iron-rich subsurface clays. There is evidence that deep-sea vents have ephemeral features that can arise and disappear on timescales of less than a decade. These structures therefore provide excellent opportunities to monitor community succession and natural selection in real time.

The diversity of marine plankton is enormous, and most of the organisms have yet to be isolated, identified and studied. If the diversity of life in the oceans is to be understood, an assessment is required of how diverse marine microscopic life is, and the driving forces of evolution in the oceans must be identified. The International Census of Marine Microbes (ICOMM; http://icomm.mbl.edu) seeks to generate an inventory of unicellular organisms, but a census is also needed of metabolism and community processes. This inventory should include all the marine microorganisms, including viruses, Bacteria, Archaea and microbial eukaryotes.

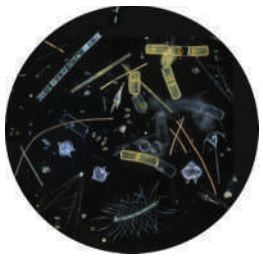
Genomics-enabled analysis of the rich diversity of microscopic life in the oceans is now possible, providing a source of information by which to decode previous life histories. The initial phase of the global ocean survey, an ambitious expedition to chart the ocean genome, generated an impressive number of open reading frames (presumed to be genes), equivalent to half of the entire GenBank inventory of known genes[8]. This study of marine bacteria highlighted the vast and previously unknown genetic information contained in extant marine microorganisms, from new protein families to novel metabolic processes. However, many of the open reading frames are unlike any known genes. They could encode metabolic processes that are yet to be discovered or be important in the regulation of cellular activity in the dynamic and variable marine environment. Whole-genome sequences from representative species of major groups are, or soon will be, available (for example, heterotrophic and photosynthetic bacteria, prasinophytes, diatoms, *Emiliania huxleyi*, *Phaeocystis* and copepods). This will not only reveal their genomic content but also provide hints about their evolutionary origin. These genomes offer a complementary understanding of diversity and complexity, and serve as anchors for interpreting ocean processes at the level of the gene.

Microarrays and probes can identify functional groups, species and ecotypes reliably and rapidly[9]. Metatranscriptomics and/or proteomics

[1]CNRS UMR8186, Department of Biology, Ecole Normale Supérieure, 46 rue d'Ulm, Paris, France. [2]Stazione Zoologica 'Anton Dohrn,' Villa Comunale, 80121 Naples, Italy. [3]School of Ocean and Earth Science and Technology, Center for Microbial Oceanography: Research and Education, University of Hawaii, Honolulu, Hawaii 96822, USA. [4]Center for Bioinformatics and Computational Biology, 3103 Biomolecular Sciences Building, University of Maryland, College Park, Maryland 20742, USA.

Box 1 | The invisible majority

Plankton are traditionally defined as marine unicellular and multicellular life forms smaller than a few millimetres (see image, courtesy of U. Sacchi and M. Montresor, Stazione Zoologica, Naples, Italy; scale bar 20 μm). This grouping by size combines organisms from all three domains of life (Bacteria, Archaea and Eukarya), despite their distinct evolutionary histories, physiological capabilities and ecological niches. Virus particles and other small, obligate parasites, although lacking a free-living existence, are also considered part of the natural microbial assembl: Their common attributes are a high rate of metabolism ar rapid generation time compared with larger organisms. They are invisible to the unaided human eye, but their metabolic capabilities and collective ecosystem service make them vital to the habitability of the planet.



Phytoplankton use solar energy and carbon dioxide to generate oxygen and the organic food that fuels higher trophic levels. Cyanobacteria are the only bacterial members of the phytoplankton; the other members are eukaryotes and comprise diatoms, dinoflagellates, coccolithophores and green algae. Phytoplankton are the main food source of zooplankton, which are composed of unicellular and multicellular organisms, as well as the juvenile stages of non-planktonic adults, which in turn are food for higher animals such as fish. Bacterioplankton are made up of certain types of Bacteria and Archaea; they are ubiquitous in the world's oceans and are the most abundant life form on our planet[38]. Just 1 litre of sea water can contain up to 1 billion bacterial cells, and viruses can be an order of magnitude more abundant. Phytoplankton and zooplankton are much less common, but population densities can increase enormously during blooms.

All microorganisms, regardless of taxonomic or physiological status, require at least three major resources to survive and proliferate: energy, electrons and carbon (and related elements, including nitrogen, phosphorus and sulphur). Depending on how these materials are obtained, microorganisms can be classified into one of three categories: photo- or chemotrophs, litho- or organotrophs, and auto- or heterotrophs[25]. For example, if solar energy is used, the microorganism is a phototroph; if chemical energy is used, it is a chemotroph. Microorganisms that do both are described as mixotrophs. In most natural habitats, there is acute competition for energy, so mixotrophy is a common metabolic strategy. The traditional scheme of autotrophy (for green plants) and heterotrophy (for all other plants) ignores the metabolic complexity of life on Earth, especially within the microbial world. The extraordinary diversity of microbial life in the ocean is due mainly to the sustained availability of energy.

Some metabolic processes occur only in selected groups of microorganisms, termed functional groups. For example, the local balance between denitrification (the removal of biologically available nitrogen) and N_2 fixation (the formation of biologically available nitrogen) can have profound impacts on ecosystem productivity. Because most open-ocean habitats are chronically short of nitrogen, the net gain of fixed nitrogen by N_2 fixation is one of the key ecological processes in the ocean.

Finally, it is well known that microorganisms assemble in a non-random fashion. These microbial assemblages are highly structured and interactive, which facilitates metabolic transformations, and gene activity is highly regulated. Symbiotic associations are also common, for example between diatoms and N_2-fixing bacteria, and self-sufficient microscopic communities can form around a single organism, such as a radiolarian or foraminifer. Some microorganisms are specialists and have streamlined genomes owing to selective gene loss over evolutionary timescales. Others are generalists and have larger and more complex genomes, but this is offset by their greater metabolic plasticity.

can elucidate metabolic activities under different conditions in the ocean in various organisms. They also allow the rapid identification of candidate genes and facilitate the association of genes with specific metabolic and regulatory functions[10] in different organisms spanning hundreds of millions of years of evolution. Metatranscriptomics can also be used to improve our genomic understanding of key organisms such as diatoms, in which recent investigations have revealed the unexpected presence of a urea cycle and hundreds of bacterial genes[11,12] (see also page 185). In this way, genome sequences from model species can be used to identify genes important in regulating ocean processes[13,14].

Looking to the future, genomic sequencing of single cells will soon become routine, and forthcoming sequencing technologies will greatly reduce the cost and extend the depth of coverage to hundreds of megabases in a single run. Metagenomics approaches will be transformed by technologies poised to appear over the next decade that will enable single-molecule sequencing up to 10 kilobases[15]. Miniaturization will allow sequencing in real time on research vessels, and the use of genomic-enabled technologies on moorings, buoys and autonomous undersea vehicles will allow studies in locations where traditional microbial methods cannot be used. The information acquired will challenge

Box 2 | The oceanic carbon cycle

The oceanic reservoir of carbon, approximately 4×10^{19} g, is distributed unequally among dissolved and particulate constituents with various chemical compositions. The largest pool is dissolved inorganic carbon (DIC), which is the most oxidized form of carbon (the valence state is +4), and the smallest pool is that comprising living organisms (mostly microorganisms), which has a much lower oxidation state (a valence state of 0 to –4). This chemical disequilibrium between oxidized and reduced carbon is produced and sustained by biological (mostly microbiological) processes. The reversible interconversion among the various forms of inorganic and organic carbon in the sea is termed the oceanic carbon cycle.

The ocean is a key component of the global carbon cycle. Approximately half of the daily photosynthetic production of organic matter on Earth takes place in the upper 100 m of the marine environment, so the oxygen in every other breath we take can be traced back to the sea. On land, large plants with long generation times (on average 10 years) are the most active contributors to photosynthesis, but in the sea it is nearly exclusively the result of rapidly growing microorganisms (with typical generation times of 1 week). Consequently, organic carbon pools in marine ecosystems are very dynamic.

The distribution of carbon in the sea is governed by two fundamentally distinct processes (termed 'pumps') that have independent controls. The 'solubility carbon pump', which transports mostly DIC, is controlled by CO_2 solubility and large-scale ocean circulation. Superimposed on these physical constraints is the less-well-understood 'biological carbon pump', which includes the production, transport and decomposition of particulate and dissolved matter, and the production and dissolution of calcium carbonate by specialized groups of organisms, including coccolithophores, foraminifers and corals. Furthermore, because particulate matter is, on average, denser than the sea water surrounding it, there is a net downward flux of carbon in the ocean resulting from gravitational settling. This process transfers energy, electrons and carbon to the deep sea and is essential for the survival of all living organisms beneath the sunlit upper portions of the water column.

The biological transfer of particulate carbon from near-surface habitats to great depths, and its subsequent decomposition and dissolution, sustains the characteristic vertical profile of DIC in the open ocean, with highest the DIC concentrations at depths of more than 1,000 m. These biological processes therefore help to sequester carbon in the deep sea, where it is stored for periods ranging from centuries to millennia. The impact of climate variability, especially greenhouse-gas-induced warming, on the efficiency of the ocean's biological carbon pump is not well understood, largely because of the complex, nonlinear behaviour of most ecological processes. Understanding this process is an important challenge for the future.

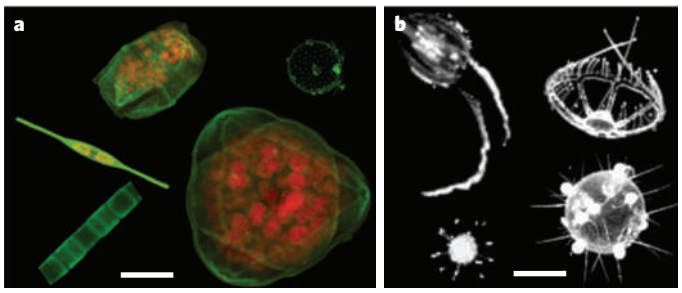


Figure 1 | Methods of visualizing plankton. **a**, Fluorescent labelling of diatoms using fluorescein-isothiocyanate-conjugated silane[37]. The dye labels all silicified structures and can be used directly on samples from the natural environment. The image shows a range of diatoms from surface waters in the Bay of Naples, Italy. Scale bar, 30 μm. (Courtesy of X. Lin and A. Amato, Ecole Normale Supérieure, Paris.) **b**, Plankton recorded *in situ* at a depth of around 500 m using the Underwater Video Profiler constructed at the Oceanography Laboratory of Villefranche-sur-Mer, France, during a cruise carried out for the California Current Ecosystem Long Term Ecological Research site. Scale bar, 1 cm. (Images courtesy of G. Gorsky and M. Picheral, Oceanography Laboratory of Villefranche-sur-Mer, France.)

historical assumptions and lead to fresh breakthroughs.

Advances in imaging technologies have revolutionized cell biology over the past 30 years. Macroscopic and microscopic imaging incorporated into oceanographic studies can connect form with function, notably showing how cell size and shape have been selected in the ocean environment (Fig. 1). Flow cytometry is already widely used in oceanography, and new imaging technologies have recently been developed for *in situ* applications[16–18]. Although the measurement of chlorophyll fluorescence has become routine, the incorporation of fluorescence microscopy into oceanographic analyses is still limited, despite advances in such technologies over recent decades. The analysis of intracellular structures in live or fixed cells by using specific fluorescent probes, revealing differences in function, such as the size of vacuoles or the extent of metabolic activity under different conditions, will be highly informative for microbial oceanography. By incorporating imaging technologies into the marine sciences, and combining them with genomic information, it will be possible to go beyond descriptive oceanography to understand more about the interactions between form, function, genotype and phenotype, and the influence of the environment on each.

The oceans and climate

One of the most serious challenges this century will be to understand how climate change — past, present and future — influences life in the oceans. We lack the adequate baseline data with which to compare contemporary observations to determine whether climate variability alters microbial metabolism and marine ecosystem services. We are in effect conducting a global-scale experiment, but with no control planet.

The oceans redistribute heat, affecting both weather and climate. Greenhouse-gas-induced temperature increases and ocean acidification are expected to have profound consequences on ocean processes. Some effects, such as increased stratification, sea-level rise and changes in ocean mixing caused by severe weather events, are already discernible[19]. Ocean acidification from increased atmospheric carbon dioxide entering the oceans is likely to affect not only calcifying organisms, such as coccolithophores and corals, but also other groups of organisms[20]. Stratification will isolate phytoplankton from the nutrients they need to capture solar energy and grow efficiently, and temperature changes are already causing species migrations over large latitudes[21]. However, the full consequences of climate change on the ocean biome are unknown because it is difficult to carry out rigorous temporal and spatial sampling and to translate laboratory or on-deck experiments to the natural environment. For example, methods such as measuring ocean colour as a proxy for carbon fixation are not ideal because chlorophyll content, biomass and photosynthetic activity are not always correlated.

Experiments that encompass wide spatial scales (from micrometres to thousands of kilometres) with appropriate temporal resolution (from seconds to millennia) have yet to be designed (Boxes 2 and 3). A further necessity is to move from the description of organisms to functional analysis, using methods that measure and monitor biological function and their ecological context (see pages 193 and 200). Genomic-enabled technologies make it possible to define functional groups by the activities of specific genes and to associate suites of gene products to a specific ocean environment (Fig. 2). Reliable biosensors for key biogeochemical processes, such as carbon fixation, nitrogen assimilation and iron bioavailability, are being developed[22–24].

A major investment has been made over several decades to collect ecological data at long-term observatories. Several ocean time-series sites exist worldwide, including locations in the open ocean[25]. Examples

Box 3 | Challenges and opportunities

Microbial oceanography is a relatively new scientific discipline that focuses on the ocean as a habitat for the evolution and regulation of microbial-based processes and their ecological consequences. It combines observation, experimentation and models, and strives to integrate the principles of several otherwise unrelated scientific disciplines[2,3]. It is truly a sea of opportunity, but a few major challenges (technical, conceptual and intellectual) preclude a comprehensive understanding of ocean processes at present. Notwithstanding the clear and urgent need for additional knowledge, the exploration of the biology of the oceans is severely under-resourced at the moment[39].

Major challenges include:
- A lack of conceptual and theoretical ecological models
- Difficulties in four-dimensional sampling of a complex and dynamic habitat because of a lack of suitable microbial and biogeochemical sensors
- Difficulties in knowing how to carry out a 'census' of marine microorganisms
- Insufficient numbers of relevant culturable model organisms
- Insufficient development of numerical simulation models for the accurate prediction of changes in microbial processes in response to climate variability
- A lack of understanding of the functional connection between human and ocean health

Responses to these challenges need to include:
- Creating and funding international collaborative research into microbial oceanography
- Recruiting and training new microbial oceanographers
- Developing new tools and methodologies for genomic-enabled oceanography
- Potentiating time-series sites globally by defining and observing functionally relevant ecosystem parameters
- Developing an international, freely accessible database that allows oceanographic and genomic data to be analysed and productively interpreted
- Establishing biologically informed definitions of functional microbial groups based on both community composition and the activities of genes, and defining organic matter using metabolomics
- Using genetically accessible model marine organisms to improve the knowledge of ecologically significant organisms and communities
- Making effective use of knowledge from non-marine model organisms
- Conducting large-scale ecosystem perturbation experiments to test hypotheses concerning microbial processes in the open sea
- Testing existing models of microbial biogeochemical processes, and revising them to evolve according to changing environmental conditions

of the latter include the Hawaii Ocean Time-series programme and the Bermuda Atlantic Time-series Study. Traditionally, physicochemical oceanographic data are collected, usually by batch collection, but more recently bacterial populations have also been monitored[10,26] using metagenomics and functional genomics. The Center for Microbial Oceanography: Research and Education in Hawaii is a recently established US National Science Foundation Science and Technology Center designed to bring about a comprehensive understanding of diverse marine planktonic assemblages. The MarMic initiative from the Max Planck Institute for Marine Microbiology in Bremen, Germany, is a similar example, focusing on marine sedimentary habitats. Such initiatives provide valuable starting points for a global holistic approach to the study of ocean dynamics. Time-series sites also serve as a 'canary in the coal mine,' providing early warning of changes. Sites of special scientific interest (such as regions where the oceans are already acidifying[27], minimal-oxygen zones[28] or locations of ice melts[29]) need to be similarly studied, and single-time-point sampling at many sites can have a complementary value by enriching baseline measurements concerning the potential range of community composition at geographically distributed sites.

Oceanographic physicochemical metadata generated at sampling sites need to be accessible in parallel with sequence data (Fig. 2). This will require the development of new database configurations accepted by the international oceanographic community. The Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) database offers a prototype[30], although the huge amount of information, for example from DNA sequencing and high-resolution imaging, presents a challenge for the future.

The oceans and human health

The balance between marine viruses and their hosts, controls on the dynamics of harmful algae, and the processes that affect nutrient concentrations in marine waters can all influence human health[31]. Destabilizing these fragile equilibria can have serious repercussions for humans and the environment. Changes in water temperature and ultraviolet radiation, two factors known to be affected by human activities, disturb the relative numbers of bacteria, fungi and viruses in the oceans, with consequences for fish and marine mammals. Fishery stocks are critical as food for human populations, especially in developing countries, and diseases caused by pathogenic microorganisms affect food availability. The use of many marine animals for food, including shellfish and many species of fish, depends on the availability of unpolluted sea water and disease-free conditions. Nutrient overloading, for example from agricultural runoff waters, provokes harmful algal blooms that are devastating for fish farms and can also poison humans and wildlife that consume contaminated shellfish. Less widely recognized is that such blooms can introduce new species that outcompete indigenous marine populations[32].

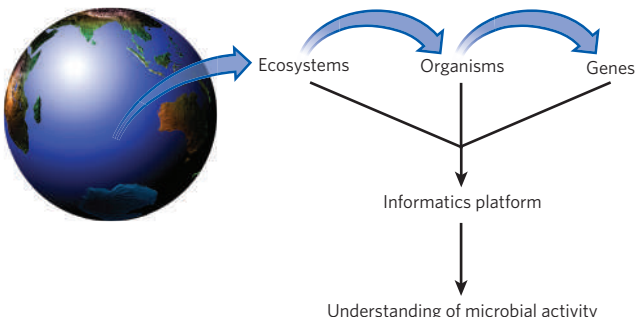


Figure 2 | Proposed framework for assessing oceanic microbial diversity in a functional context. Three levels of data should be collected from each system under study: ecosystem physicochemical data, composition of organisms, and expressed genes. The contextualization of these three sorts of data by dedicated informatics platforms should allow an understanding of microbial activity with respect to the prevailing ocean conditions.

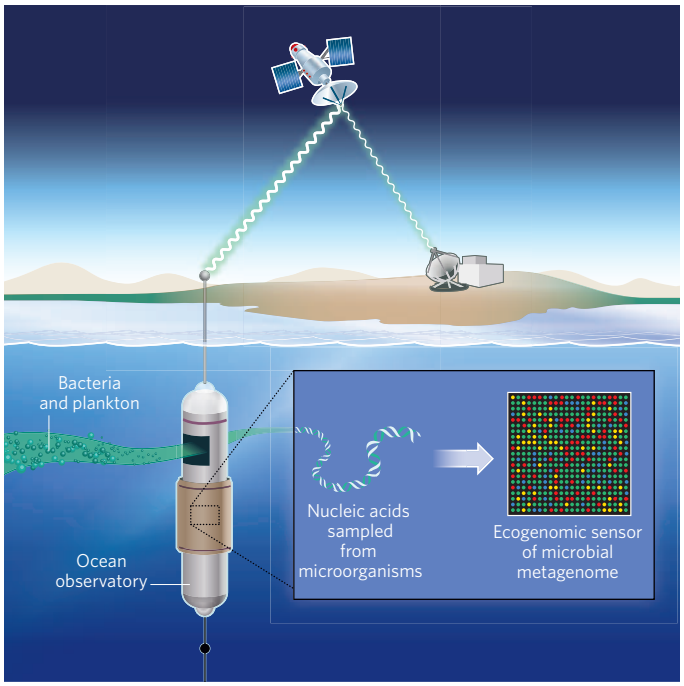


Figure 3 | Miniaturized ecogenomic sensors to measure microbial activity. The sensors could be installed into advanced ocean observatories to monitor DNA and RNA from diverse microbial communities. Subsystems for monitoring, data management and communication, and data modelling would be incorporated for data contextualization. The sensors would report to a worldwide network of laboratories in real time by satellite telemetry.

Coastal zones of the world's oceans are increasingly subjected to the discharge of human waste products, ranging from domestic to industrial effluents. The result is a loss of seagrasses and related estuarine and marine vegetation and the build-up of bacteria and viruses with pathogenic potential. Recreational areas along the coasts become both public health hazards and an aesthetic loss for communities. The extent of plastic debris in several open-ocean regions worldwide, notably the Great Pacific Garbage Patch, is a major threat to ocean life[33].

The human pathogen *Vibrio cholerae*, which causes cholera, is native to coastal and estuarine environments. Coastal temperature is a key determinant of the burden of cholera in coastal waters[34]. Conversely, the life history of this bacterium in Bangladesh is closely linked to the occurrence of planktonic blooms in the Bay of Bengal months before the outbreak of disease[35]. The role of *V. cholerae* in the ecology of the marine environment is extensive: the bacterium can digest chitin, degrade petroleum and carry out denitrification. The links, in evolutionary terms, between human and animal pathogens and their non-pathogenic marine relatives are only now beginning to be analysed[36]. The study of cholera epidemics in human populations represents a useful case study to improve our understanding and prediction of human disease outbreaks. Furthermore, cholera is similar to many other vector-borne human diseases, such as malaria and dengue fever, in being highly sensitive to climate.

In addition to understanding how the oceans affect human health, as a result of infectious disease and ocean pollution, we will need to learn how rising sea levels and altered ocean circulation affect the distribution of microbial populations, because both are affected by human activities and climate events[2]. Marine systems are more highly interconnected than terrestrial systems, so an alteration in microbial equilibria in one part of the ocean can affect a geographically remote area.

Looking forward

Contemporary oceanography, enabled by microbial genomics and other modern technologies, represents a maturation of descriptive biological oceanography. We need to define functional 'keystone' groups by their

unique genomic signatures that can be measured and compared across contrasting ecosystems. Traditional nutrient classification schemes, such as dissolved organic matter, will need to be redefined within an ocean metabolome that incorporates knowledge of the metabolic potential of individual components for different species and communities (Box 3). Model microbial systems and laboratory-based experimentation combined with open-ocean observation (Fig. 3) will give us a holistic perspective, moving beyond reductionist science to link descriptive and functional observations, improving the predictive power of oceanography. Realizing such ambitious goals at a time of accelerating global climate change will depend on international collaboration, both to meet these challenges and to educate the next generation of oceanographers. Success will require cross-disciplinary collaborations encompassing and integrating multi-hierarchical and multi-scalar measurements and incorporating databases from each discipline into an inter-operational database. The ultimate goal is the orchestration of a grand synthesis of emergent models that transcends each of its component parts. ■

1. Curtin, T. B. & Belcher, E. O. Innovation in oceanographic instrumentation. *Oceanography* **21,** 44–53 (2008).
2. Dawson, M. N., Sen Gupta, A. & England, M. H. Coupled biophysical global ocean model and molecular genetic analyses identify multiple introductions of cryptogenic species. *Proc. Natl Acad. Sci. USA* **102,** 11968–11973 (2005).
3. Follows, M. J., Dutkiewicz, S., Grant, S. & Chisholm, S. W. Emergent biogeography of microbial communities in a model ocean. *Science* **315,** 1843–1846 (2007).
4. Falkowski, P. G., Fenchel, T. & Delong, E .F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320,** 1034–1039 (2008).
5. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281,** 237–240 (1998).
6. Kolber, Z. Energy cycle in the ocean: powering the microbial world. *Oceanography* **20,** 79–87 (2007).
7. Longhurst, A. R. *Ecological Geography of the Sea* (Academic, 1998).
8. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** e77 (2007).
9. Huber, J. A. *et al.* Microbial population structures in the deep marine biosphere. *Science* **318,** 97–100 (2007).
10. Frias-Lopez, J. *et al.* Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA* **105,** 3805–3810 (2008).
11. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306,** 79–86 (2004).
12. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456,** 239–244 (2008).
13. Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105,** 10438–10443 (2008).
14. Mock, T. *et al.* Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proc. Natl Acad. Sci. USA* **105,** 1579–1584 (2008).
15. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).
16. Brandt, A. *et al.* First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature* **447,** 307–311 (2007).
17. Olson, R. J. & Sosik, H. M. A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot. *Limnol. Oceanogr. Methods* **5,** 195–203 (2007).
18. Stemmann, L. *et al.* Global zoogeography of fragile macrozooplankton in the upper 100–1000 m inferred from the underwater video profiler. *ICES J. Mar. Sci.* **65,** 433–442 (2008).
19. Polovina, J. J., Howell, E. A. & Abecassis, M. Ocean's least productive waters are expanding. *Geophys. Res. Lett.* **35,** L03618 (2008).
20. Hutchins, D. A. *et al.* CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnol. Oceanogr.* **52,** 1293–1304 (2007).
21. Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A. & Edwards, M. Reorganization of North Atlantic marine copepod biodiversity and climate. *Science* **296,** 1692–1694 (2002).
22. John, D. E. *et al.* Phytoplankton carbon fixation gene (RuBisCO) transcripts and air–sea CO_2 flux in the Mississippi River plume. *ISME J.* **1,** 517–531 (2007).
23. Prasil, O., Suggett, D. J., Cullen, J. J., Babin, M. & Govindjee. Aquafluo 2007: chlorophyll fluorescence in aquatic sciences, an international conference held in Nové Hrady. *Photosynth. Res.* **95,** 111–115 (2008).
24. Short, S. M. & Zehr, J. P. Nitrogenase gene expression in the Chesapeake Bay Estuary. *Environ. Microbiol.* **9,** 1591–1596 (2007).
25. Karl, D. M. Microbial oceanography: paradigms, processes and promise. *Nature Rev. Microbiol.* **5,** 759–769 (2007).
26. Delong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311,** 496–503 (2006).
27. Feely, R. A., Sabine, C. L., Hernandez-Ayon, J. M., Ianson, D. & Hales, B. Evidence for upwelling of corrosive 'acidified' water onto the continental shelf. *Science* **320,** 1490–1492 (2008).
28. Stramma, L., Johnson, G. C., Sprintfall, J. & Mohrholz, V. Expanding oxygen-minimum zones in the tropical oceans. *Science* **320,** 655–658 (2008).
29. Murray, A. E. & Grzymski, J. J. Diversity and genomics of Antarctic marine micro-organisms. *Phil. Trans. R. Soc. Lond. B* **362,** 2259–2271 (2007).
30. Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P. & Frazier, M. CAMERA: a community resource for metagenomics. *PLoS Biol.* **5,** e75 (2007).
31. Colwell, R. R. *et al.* Climate change and human health. *Science* **279,** 968–969 (1998).
32. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321,** 926–929 (2008).
33. Thompson, R. C. *et al.* Lost at sea: where is all the plastic? *Science* **304,** 838 (2004).
34. Constantin de Magny, G. *et al.* Environmental signatures associated with cholera epidemics. *Proc. Natl Acad. Sci. USA* **105,** 17676–17681 (2008).
35. Alam, M. *et al.* Viable but nonculturable *Vibrio cholerae* O1 in biofilms in the aquatic environment and their role in cholera transmission. *Proc. Natl Acad. Sci. USA* **104,** 17801–17806 (2007).
36. Nakagawa, S. *et al.* Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl Acad. Sci. USA* **104,** 12146–12150 (2007).
37. Desclés, J. *et al.* New tools for labeling silica in living diatoms. *New Phytol.* **177,** 822–829 (2008).
38. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
39. Richardson, A. J. & Polooczanska, E. S. Under-resourced, under threat. *Science* **320,** 1294–1295 (2008).

Acknowledgements Many of the themes discussed here were inspired by a workshop entitled 'Implications and Opportunities of the Marine Genomics Revolution' held in Bermuda in October 2007 and sponsored by the Marine Genomics Europe Network of Excellence (GOCE-CT-2004-505403), the Centre National de la Recherche Scientifique (CNRS), the Gordon and Betty Moore Foundation (GBMF), the US National Science Foundation (NSF), the US Office of Naval Research Global, the Bermuda Institute of Ocean Sciences, and the Japanese Science and Technology Agency. We thank the participants of this meeting as well as the anonymous reviewers for helpful comments on the manuscript. We apologize to colleagues whose work we could not cite because of space restrictions. Work in our laboratories is supported by the European Union, the Agence Nationale de la Recherche and CNRS (C.B.), the NSF (grant number EF-04245599), GBMF and Agouron Institute (D.M.K.), and the NSF (grant number 0813066), the National Institutes of Health (grant number 1RO1A139129-01) and the National Oceanic and Atmospheric Administration (grant number S0660009) (R.R.C.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to C.B. (cbowler@biologie.ens.fr).

The life of diatoms in the world's oceans

E. Virginia Armbrust[1]

Marine diatoms rose to prominence about 100 million years ago and today generate most of the organic matter that serves as food for life in the sea. They exist in a dilute world where compounds essential for growth are recycled and shared, and they greatly influence global climate, atmospheric carbon dioxide concentration and marine ecosystem function. How these essential organisms will respond to the rapidly changing conditions in today's oceans is critical for the health of the environment and is being uncovered by studies of their genomes.

About one-fifth of the photosynthesis on Earth is carried out by microscopic, eukaryotic phytoplankton known as diatoms[1]. These photosynthetic workhorses are found in waters worldwide, wherever there is sufficient light and nutrients. Their name is derived from the Greek *diatomos*, meaning 'cut in half', a reference to their distinctive two-part cell walls made of silica (Fig. 1). Each year, diatom photosynthesis in the sea generates about as much organic carbon as all the terrestrial rainforests combined[1,2]. But unlike much of the carbon generated by trees, the organic carbon produced by diatoms is consumed rapidly and serves as a base for marine food webs. In coastal waters, diatoms support our most productive fisheries. In the open ocean, a relatively large proportion of diatom organic matter sinks rapidly from the surface, becoming food for deep-water organisms[3]. A small fraction of this sinking organic matter escapes consumption and settles on the sea floor, where it is sequestered over geological timescales in sediments and rocks and contributes to petroleum reserves. Given the crucial role of diatoms in the global carbon cycle, plans have been made, controversially, to reduce atmospheric levels of the greenhouse gas carbon dioxide by fertilizing large regions of the ocean with iron to generate huge blooms of diatoms[4].

Fresh insight into the mechanisms underlying the global impact of diatoms came from the availability of the roughly 34 megabases (Mb) of DNA sequence for the nuclear, plastid and mitochondrial genomes of the cosmopolitan diatom *Thalassiosira pseudonana*[5]. Whole-genome sequence of a second model diatom, *Phaeodactylum tricornutum* (27 Mb), soon followed[6], and draft sequence is now available for the polar species *Fragilariopsis cylindrus* (80 Mb) and the toxigenic coastal species *Pseudo-nitzschia multiseries* (300 Mb). One of the more intriguing outcomes of the sequencing projects thus far is a recognition of the unique combination of genes and metabolic pathways that distinguish diatoms from the evolutionarily distinct plant and animal lineages. Enormous amounts of diversity are encapsulated within diatoms. For example, *T. pseudonana* and *P. tricornutum* probably diverged from one another only about 90 million years (Myr) ago, yet their genomes are about as different as those of mammals and fish, which diverged about 550 Myr ago[6].

Here I explore the intersection of diatom ecology, biogeochemistry and genomics, with a focus on the roles of diatoms in past and contemporary oceans. What emerges is a genomics-based reflection of the complex interactions that define marine ecosystems, in which metabolites and capabilities are shared across different kingdoms of organisms. The goal is twofold: first, to provide a window into the fascinating world of this unusual group of organisms that has such a crucial role in regulating the stability of our planet; and second, to gain a deeper understanding of how diatoms may fare under future ocean conditions. This is crucial because alterations in diatom populations resulting from climate change could have a dramatic effect on Earth's atmosphere.

Life in the ocean waves

Marine microbial communities are incredibly diverse, consisting of interconnected groups of cyanobacteria, heterotrophic Bacteria, Archaea, viruses, eukaryotic phytoplankton and protists. The most abundant phytoplankton in the sea are the marine cyanobacteria of the genus *Prochlorococcus*. The most diverse group of phytoplankton is the diatoms, with an estimated 200,000 different species, ranging in size from a few micrometres to a few millimetres and existing either as single cells or as chains of connected cells[7] (Fig. 1). Diatoms reproduce primarily by mitotic divisions interrupted infrequently by sexual events (Box 1). They bloom quickly, increasing in cell number by many orders of magnitude in just a few days. Diatoms tend to dominate phytoplankton communities in well-mixed coastal and upwelling regions, as well as along the sea-ice edge, where sufficient light, inorganic nitrogen, phosphorus, silicon and trace elements are available to sustain their growth[8]. In polar environments, where glaciers and permafrost limit photosynthesis on land, diatoms are critical components of the food webs that sustain both marine and terrestrial ecosystems. Larger species of diatoms can move up and down through the water column by controlling their buoyancy. Certain open-ocean species can move between well-lit but nutrient-depleted surface waters, in which they photosynthesize, and nitrate-rich waters at a depth of about 100 m, where they take up and store the nutrients necessary to keep dividing[9]. Diatoms seem to have exquisite communication capabilities, using a nitric-oxide-based system that mediates signalling between and within cells and regulates the production of aldehydes[10], which can be harmful to grazing copepods[11].

Mix-and-match genomes

Diatoms have a complex evolutionary history that is distinct from plants, the dominant photosynthetic organisms on land[12]. Oxygenic photosynthesis had its origins in cyanobacteria, but different endosymbiotic events gave rise to plants and diatoms (Fig. 2). The initial, primary, endosymbiosis occurred about 1.5 billion years ago, when a eukaryotic heterotroph engulfed (or was invaded by) a cyanobacterium to form the photosynthetic plastids of the Plantae, the group that includes land plants and red and green algae[13]. Genes were subsequently transferred from the symbiotic cyanobacterial genome to the host nucleus, with about 10% of Plantae nuclear genes being derived from the cyanobacterial endosymbiont[14]. About 500 million years later, a secondary endosymbiosis occurred, in which a different eukaryotic heterotroph captured a red alga. Over time, the red-algal

[1]School of Oceanography, University of Washington, Seattle, Washington 98195, USA.

endosymbiont was transformed into the plastids of the Stramenopiles, the group that includes diatoms, brown macroalgae and plant parasites. Gene transfer continued from the red-algal nuclear and plastid genomes to the host nucleus[5]. At least 170 red-algal genes have been identified in the nuclear genome of diatoms, most of which seem to encode plastid components[6]. As in Plantae plastids, photosynthesis and the biosynthesis of fatty acids, isoprenoids and amino acids are carried out in diatom plastids.

Diatoms have a distinctive range of attributes that can be traced to this union between heterotrophic host and photosynthetic red alga. For example, unlike plants, diatoms have a complete urea cycle, although it remains to be seen how they use this pathway. The urea cycle was previously thought to be restricted to organisms that consume complex organic nitrogen compounds and excrete nitrogenous waste products[5]. Diatoms also combine an animal-like ability to generate chemical energy from the breakdown of fat with a plant-like ability to generate metabolic intermediates from the breakdown, a combination that probably allows diatoms to survive long periods of darkness, as occurs at the poles, and resume division and growth when they return to the light[5]. Numerous examples of this mix-and-match compilation of characteristics reiterate the simple fact that diatoms are neither plants nor animals.

More recent analyses suggest additional contributions to the mixture that defines diatom genomes. One unexpected twist was revealed by comparative analyses with the Chlamydiae, a group of intracellular bacteria that today exist only as pathogens or symbionts. The presence of some chlamydial genes in both plants and red algae, but not in cyanobacteria, suggests that a chlamydial endosymbiont also tagged along during the early stages of the primary endosymbiosis[15]. Further analysis suggests that in addition to the red alga, a green alga may have contributed to the mix of nuclear genes in diatoms[16].

A second twist is the finding that at least 587 genes in the *P. tricornutum* nuclear genome seem to share a history with diverse lineages of bacteria in addition to the Chlamydiae[6]. Some bacterial genes replaced homologous genes found in other phototrophs, whereas others provided new functions to the diatoms[17]. Less than half the bacterial genes in *P. tricornutum* are shared with *T. pseudonana*, and only 10% are shared between *T. pseudonana* and the distantly related oomycete *Phytophthora*[6], suggesting that independent gains and losses of bacterial genes by diatoms are ongoing. Finally, viruses also seem to mediate gene transfer to diatoms[18], although the extent of this process is still unclear[12].

The emerging picture is that the different species of diatoms are characterized by a complex combination of genes and pathways acquired from a variety of sources (Fig. 2). Endosymbiotic events defined the overall capabilities of diatoms, but subsequent gains (or losses) of specific genes, largely from bacteria, presumably helped them adapt to new ecological niches. What factors might underlie this constantly evolving mixture of attributes? Bacteria in the sea outnumber diatoms by many orders of magnitude, ensuring that diatoms are never free of bacterial influences. There are numerous examples of diatom dependency on bacterial metabolites such as vitamins[19,20] and of bacterial dependency on released diatom metabolites[21,22]. Some bacteria attach to diatoms by embedding themselves in the crevices of diatom cell walls[23]. Open-ocean diatoms can harbour nitrogen-fixing cyanobacteria under their silica cell wall, whereas other nitrogen-fixing cyanobacteria attach to silica spines protruding from the walls[24]. In the most extreme example, bacteria have been reported to live between the third and fourth outermost membranes of the plastids of a freshwater diatom[25]. All this was possible because diatoms evolved in a dilute world where essential metabolites are shared across kingdoms. Redundancy, reliability and ease of transfer of different components of this metabolic soup in different environments probably influence whether the cross-kingdom interaction is opportunistic or an obligate symbiosis, perhaps with the incorporation of critical bacterial genes as an end point. Studying diatoms in the sterile environment of the laboratory is an important first step in predicting responses to environmental change, but new sequencing technologies that yield greater amounts of information at a lower cost provide opportunities to study these organisms in laboratory consortia that may more closely mimic the real world. Ultimately, metagenomic tools amenable to studying organisms with large genomes will be essential for understanding how diatoms function in nature.

The rise of diatoms

Molecular-clock-based estimates suggest that diatoms arose in the Triassic period, perhaps as early as 250 Myr ago[26], although the earliest well-preserved diatom fossils come from the Early Jurassic, some 190 Myr ago[27]. Before the diatoms, the phytoplankton consisted primarily of cyanobacteria and green algae only slightly larger than bacteria[28]. The emergence of diatoms and two other groups of larger eukaryotic phytoplankton, the dinoflagellates and coccolithophorids, resulted in a major shift in global organic carbon cycling. This initiated an era of declining atmospheric CO_2 concentrations and increasing atmospheric

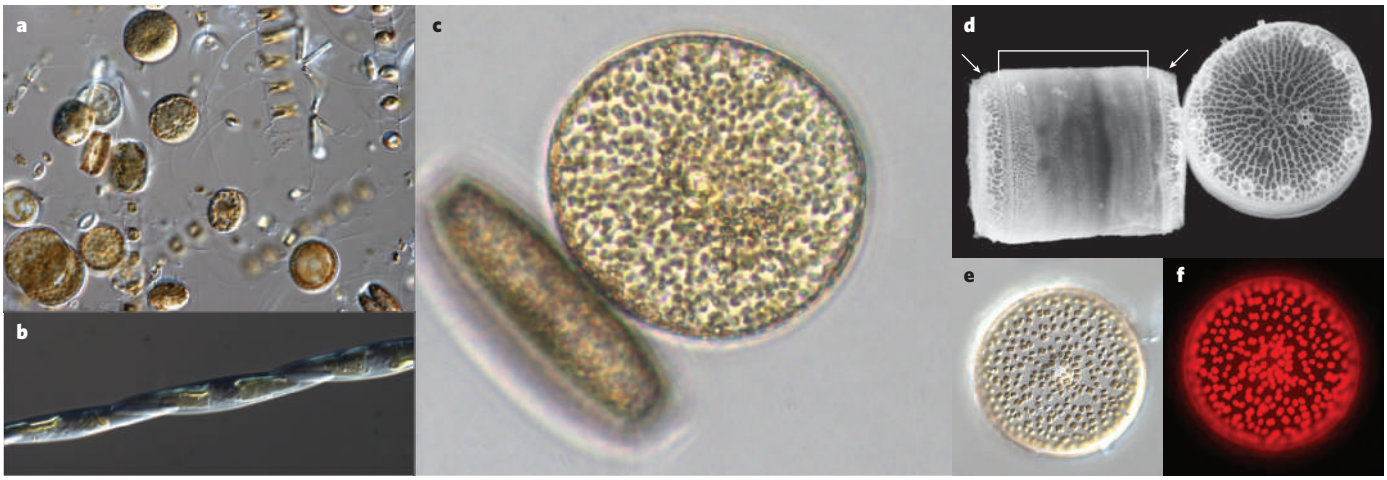
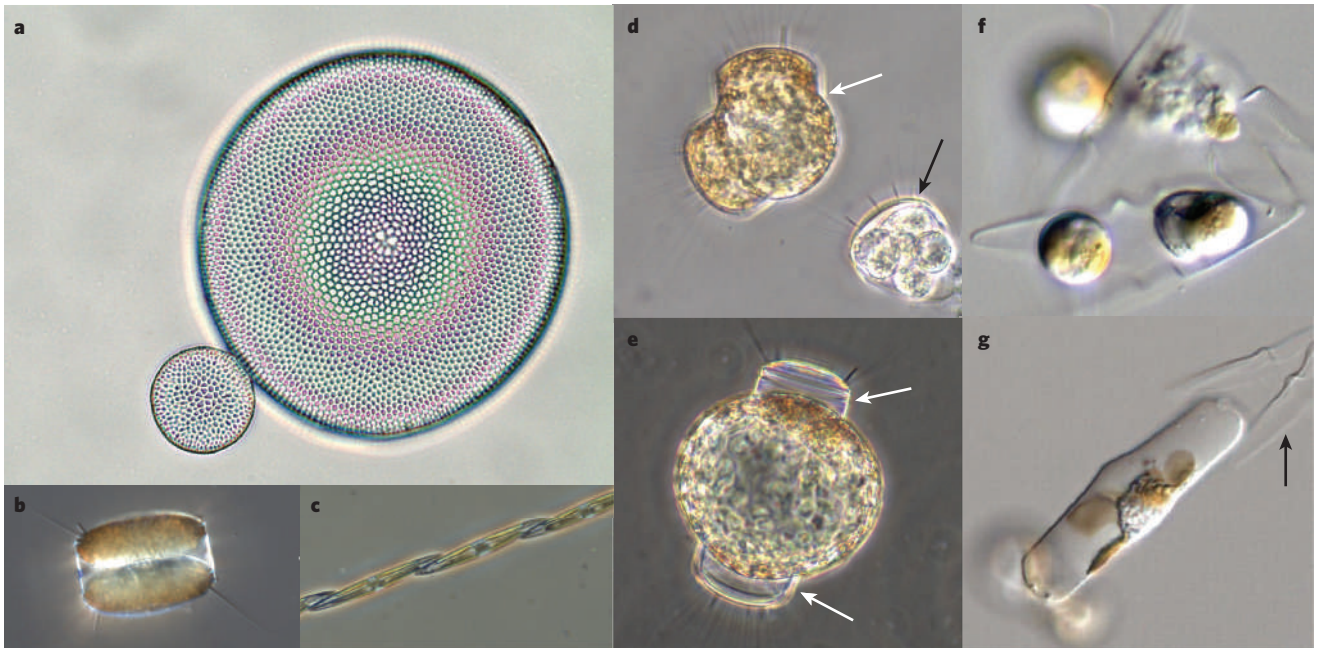


Figure 1 | Micrographs of different diatom species. **a**–**c**, Diatoms can exist as single cells or chains of cells, as illustrated in a concentrated field sample (**a**). The two main morphological categories of diatoms are pennate (**b**, *Pseudo-nitzschia*) and centric (**c**, *Thalassiosira*). The two halves of the cell wall (valves) fit together like a Petri dish (**c**) that appears round (or oval) when viewed from the top and rectangular when viewed from the side. **d**, The two valves (arrows) are held together by a series of siliceous hoops, or girdle bands (brackets), seen in more detail in this scanning electron micrograph of *Thalassiosira pseudonana*. New girdle bands are laid down during cell growth. The patterns of pores and other cell-wall structures are species specific. **e**, **f**, Photosynthesis takes place in membrane-bound plastids that appear as small discs within the cell (**e**) and contain the photosynthetic pigment chlorophyll *a*, which fluoresces red when illuminated with blue light (**f**). (Images courtesy of K. Holtermann, University of Washington, Seattle (**a**, **c**, **e**, **f**), P. von Dassow, Station de Biologique de Roscoff, France (**b**) and N. Kröger, Georgia Institute of Technology, Atlanta (**d**).)

Box 1 | The sex life of diatoms



Living in a glass box has created some interesting life-cycle attributes for diatoms. Physical and developmental constraints associated with replication of the cell wall mean that in each mitotic division, one daughter cell is slightly smaller than the other. Over successive divisions, cells of dramatically different sizes are found within a population (**a**, *Coscinodiscus*). Cell size is ultimately restored through sexual reproduction, which occurs differently in centric (**b**) and pennate (**c**) diatoms. In centric diatoms (**b**, *Thalassiosira*), only small cells are receptive to an environmental trigger and can become either sperm (**d**, white arrow), which break free of the wall, or eggs, which remain encased within the wall (**d**, black arrow). Sperm swim to an egg, gain entry past the glass wall, and fertilize the egg nucleus. The resultant zygote swells to form a specialized cell known as the auxospore (**e**), sheds its old cell walls (**e**, arrows) and produces a much larger wall, restoring cell size. This is risky for a centric diatom because if sperm are unable to find eggs in the dilute ocean, the gametes will die. Pennates (**c**, *Pseudo-nitzschia*) also have a size requirement for the initiation of sexual reproduction, but seem to form gametes only when they find an appropriate mate of the opposite sex, a seemingly less risky option. When paired, pennate cells produce morphologically identical gametes (**f**), which are unable to swim and instead move towards one another in an amoeba-like fashion and fuse to create the zygote and auxospore (**g**), which breaks free of the old cell wall (arrow). The sexual cycle of most diatoms cannot be controlled in the laboratory, hindering the development of classical genetic studies. Instead, genetic manipulation of diatoms has relied primarily on the addition of new versions of genes (transformation) or on reduced expression of targeted genes (RNA interference)[12]. (Images courtesy of J. Koester, University of Washington, Seattle (**a**), P. von Dassow, Station Biologique de Roscoff, France (**b**, **d**, **e**) and K. Holtermann, University of Washington, (**c**, **f**, **g**).)

O_2 concentrations[29]. The increased sinking rates associated with these large phytoplankton led to enhanced burial of organic carbon in continental margins and shallow seas, creating most of the petroleum reserves known today[30]. Consider that the early Mesozoic ocean differed dramatically from modern oceans. Atmospheric concentrations of CO_2 were almost eight times higher than today[31]; the average global temperature was significantly higher; and Africa and Europe were beginning to separate, leading to extensive flooding of continental shelves. This probably led to increased continental weathering and released large amounts of nutrients, which increased phytoplankton activity[29,32]. Additionally, the absence of polar ice caps and the smaller pole-to-equator temperature gradient reduced ocean circulation and increased stratification of the water column. Together, these factors decreased the oxygenation of the oceans and contributed to ocean anoxic events. Fossil records reveal that the earliest diatoms, known as radial centrics (Fig. 3), had a heavy, highly silicified cell wall that initially restricted them to a benthic lifestyle in shallow, near-shore waters.

Diatoms assumed their dominant role in the carbon cycle about 100 Myr ago, during the Cretaceous period, when atmospheric CO_2 levels were still about five times higher than they are today[31] and O_2 levels were increasing[30]. Ocean stratification was decreasing as the nutrient supply to surface waters increased. The proliferation of diatoms and other photosynthetic organisms during this period increased the oxygenation of surface waters with a concomitant decrease in iron availability. These conditions coincided with the divergence of a second major lineage of diatoms, the bipolar and multipolar centrics, which includes members of *Thalassiosira,* the genus first chosen for whole-genome sequencing (Fig. 3).

The mass extinction at the end of the Cretaceous, about 65 Myr ago, led to loss of about 85% of all species, including substantial reductions in the diversity of marine dinoflagellates and coccolithophorids[28]. Diatoms survived this event relatively unscathed and began to colonize offshore areas, including the open ocean. Centric species that migrated into the open ocean were able to survive despite the reduced nutrient levels there, further increasing their impact on the global carbon cycle. A third group of diatoms, the araphid pennates, are detected in the fossil record from this period (Fig. 3).

By 50 Myr ago, atmospheric O_2 concentrations had stabilized to today's levels, further reducing iron concentrations in the open ocean, and atmospheric CO_2 concentrations continued to decline to near today's levels[31]. Diatom diversity peaked at the Eocene/Oligocene boundary, some 30 Myr ago[33], and a fourth group of diatoms, the raphid pennates, emerged (Fig. 3). These are distinguished by a slit (raphe) in their walls that allows them to glide along surfaces[7,27]. The evolution of the raphe greatly expanded the ecological niches available and probably had as profound an impact on diatom diversification as the evolution of flight had on birds[27]. Today, the raphid species *Fragilariopsis kerguelensis* dominates the diatom community in the Southern Ocean, the largest region of diatom-based carbon export[34]. Three of the four diatoms with complete or draft genome sequences are raphid pennates. Genome-based detection of differences in the regulation of

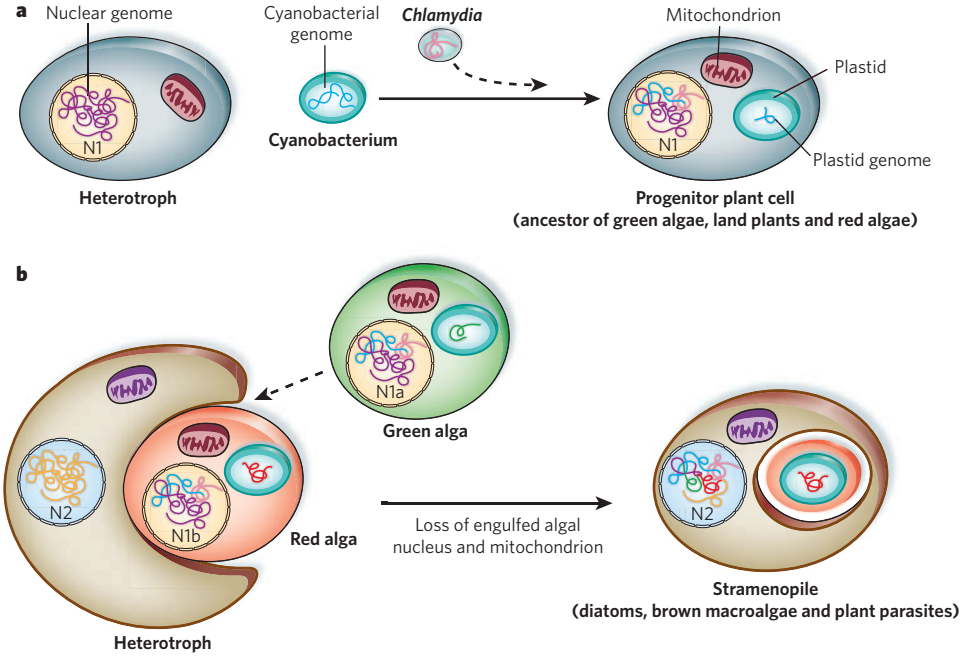


Figure 2 | Endosymbiosis in diatoms. Representation of the origin of diatom plastids through sequential primary (**a**) and secondary (**b**) endosymbioses, and their potential effects on genome evolution. **a**, During primary endosymbiosis, a large proportion of the engulfed cyanobacterial genome is transferred to the host nucleus (N1), with few of the original genes retained within the plastid genome. The potential for invasion of the host by a chlamydial parasite is indicated with a dashed arrow, and the ensuing transfer of chlamydial genes to the host nucleus is indicated in pink. The progenitor plant cell subsequently diverged into red and green algae and land plants, readily distinguished by their plastid genomes. **b**, During secondary endosymbiosis, a different heterotroph engulfs a eukaryotic red alga. Potential engulfment of a green algal cell as well is indicated with a dashed arrow. The algal mitochondrion and nucleus are lost, and crucial algal nuclear and plastid genes (indicated in blue, purple and pink) are transferred to the heterotrophic host nucleus, N2. Additional bacterial genes are gained and lost throughout diatom evolution, but for simplicity this is not indicated here.

carbon fixation by the bipolar centric *T. pseudonana* and the raphid pennate *P. tricornutum*[35] may reflect the different atmospheric CO_2 and O_2 conditions, and the resultant seawater chemistries, when the two lineages emerged.

The need for iron

Primary productivity in 30–40% of the world's contemporary oceans is limited by the availability of iron, particularly in open-ocean regions of the Southern Ocean, equatorial Pacific Ocean and north Pacific Ocean[36]. These high-nutrient, low-chlorophyll (HNLC) regions are characterized by exceedingly low concentrations of iron and high concentrations of other essential nutrients, such as nitrate, phosphate and silicic acid. Diatoms in the open ocean reduce their iron requirements under iron-limiting conditions[37,38]. Open-ocean centric species of *Thalassiosira*, for example, seem to have permanently modified their photosynthetic apparatus to require less iron[39] and have replaced iron-requiring electron-transport proteins with equivalent ones that need copper[40]. These changes seem to have compromised their ability to deal with the rapidly fluctuating light fields more characteristic of coastal environments[39]. Raphid pennate diatoms can also greatly reduce their iron requirements[41], but they seem to do so using more flexible modifications, avoiding irreversible compromises. When starved of iron, *P. tricornutum* downregulates processes that require a lot of iron, such as photosynthesis, mitochondrial electron transport and nitrate assimilation[42]. By way of compensation, these iron-limited cells restructure their proteome, upregulate alternative pathways for dealing with oxidative stress, and upregulate additional iron-acquisition pathways[42]. The presence of different iron-responsive genes suggests that raphid pennate and bipolar centric diatoms have fundamentally different systems for acquiring iron[41,42].

Centric and pennate diatoms also differ in their ability to store iron, a critical attribute for existence in the open ocean, where the iron supply is sporadic[43]. Members of the raphid pennate genera *Pseudo-nitzschia*, *Fragilariopsis* and *Phaeodactylum* all produce ferritin, an iron-storage molecule that also protects against oxidative stress[44]. No other members of the Stramenopiles seem to encode ferritin, including *T. pseudonana*, and it seems that this gene may have arisen in the restricted subset of pennate diatoms through a lateral gene transfer from another organism[44]. The enhanced iron storage provided by ferritin in *Pseudo-nitzschia* probably underlies its numerical dominance in the massive diatom blooms that result from iron fertilization (Fig. 4) and helps to explain the importance of raphid diatoms in regulating the flux of CO_2 into surface waters[44]. So far, the *T. pseudonana* genome has provided no clues to how centric diatoms store iron, suggesting that novel mechanisms are used.

Living in glass houses

One of the most striking features of diatoms is their beautiful cell wall made essentially of hydrated glass ($SiO_2.nH_2O$)[45] (Fig. 1 and Box 1). In creating these walls from silicon dissolved in sea water as silicic acid, diatoms control the biogenic cycling of silicon in the world's oceans to such an extent that every atom of silicon entering the ocean is incorporated into a diatom cell wall on average 39 times before being buried on the sea floor[46]. Depending on the conditions[1], cell walls from dead diatoms can accumulate on the sea floor as immense deposits of silica up to 1,400 metres thick, as found on Seymour Island in the eastern Antarctic Peninsula[27]. The resulting diatomaceous earth has a variety of uses, including as flea powder, insulation and toothpaste ingredients.

The elaborate species-specific patterns of nano-scale to micro-scale pores, ridges and tubular structures are genetically controlled, although external factors such as salinity influence the density and pore size of the precipitated silica[47]. Their ability to produce silica structures in three dimensions has made diatoms attractive models for nanotechnology[45] and has prompted extensive searches for components of the necessary genetic machinery. The cell wall is produced in an acidic silica-deposition vesicle and encased in an organic matrix that is rich in proteins and

sugars, preventing the silica from dissolving in sea water. Consumption of this matrix by bacteria accelerates the recycling of silicon within surface waters[48]. Three categories of molecule normally embedded directly within the wall can precipitate silica in artificial systems: silaffins, which are highly modified phosphoproteins[49]; long-chain polyamines[50]; and silacidins, which are acidic proteins[51]. Both the amount and structure of each type of molecule differ between species, consistent with them having a critical role in the species-specific patterns of cell-wall nanostructures. More than 150 additional gene products potentially required for silicon biomanipulation have been identified in *T. pseudonana*[52]. Half of the genes are upregulated when cells are starved of either silicon or iron, suggesting that the iron and silicon pathways are linked. A similar connection between iron and silicon pathways was not reported for *Phaeodactylum*[42], although it is not yet clear whether this reflects another difference between centric and pennate diatoms or the fact that *Phaeodactylum* is the only known diatom that lacks an obligate requirement for silicon.

Interactions between iron availability and silicon usage by diatoms in the Southern Ocean are thought to explain, in part, the reduction in atmospheric concentrations of CO_2 during glacial periods[53], when iron concentrations were higher. Under iron-rich conditions, diatom communities use less silicon relative to nitrogen, leaving excess silicate in surface waters. This excess then circulates out of the Southern Ocean and fuels diatom, rather than coccolithophorid, productivity in the subtropics[54]. During glacial intervals, the increased amount of organic matter produced by diatoms is thought to sink into deep waters, resulting in long-term sequestration of atmospheric CO_2 (ref. 54).

Several studies have mimicked these glacial-interval effects of iron on diatoms by fertilizing iron-limited regions of the ocean with iron. The hope of some is that this will increase the export of carbon from surface waters and thus slow the rising levels of atmospheric CO_2 generated by the burning of fossil fuel. Iron-enrichment experiments done so far confirm that iron fertilization does produce the expected diatom blooms (Fig. 4), but most of the organic carbon generated by the bloom is consumed and recycled in surface waters. There is a relatively small increase in the amount that sinks to deep waters[55]. Even large-scale fertilization projects can be expected to draw down just a small fraction of the accelerating amounts of CO_2 entering the atmosphere, and even this has the potential to shift community composition and generate other greenhouse gases[4]. If we want to sequester large amounts of CO_2, we must look elsewhere for a solution.

Deadly diatoms

The ability of diatoms to affect humans is not limited to their role in the global carbon cycle. In 1987, 107 people became ill and 3 people died after eating mussels contaminated with the powerful neurotoxin domoic acid. Detective work showed that the toxin was produced by the diatom *P. multiseries*[56], whose genome is currently being sequenced. Domoic acid is water soluble and binds to glutamate receptors, causing a massive depolarization of nerve cells, particularly in the hippocampus. Crustaceans, bivalves and fish all serve as vectors of domoic acid to humans and other vertebrates[56]. Careful monitoring of shellfish has prevented further documented incidents in humans, but wild animals, including birds and mammals, are increasingly being affected by domoic acid[57].

The genus *Pseudo-nitzschia*, which is the chief culprit, comprises at

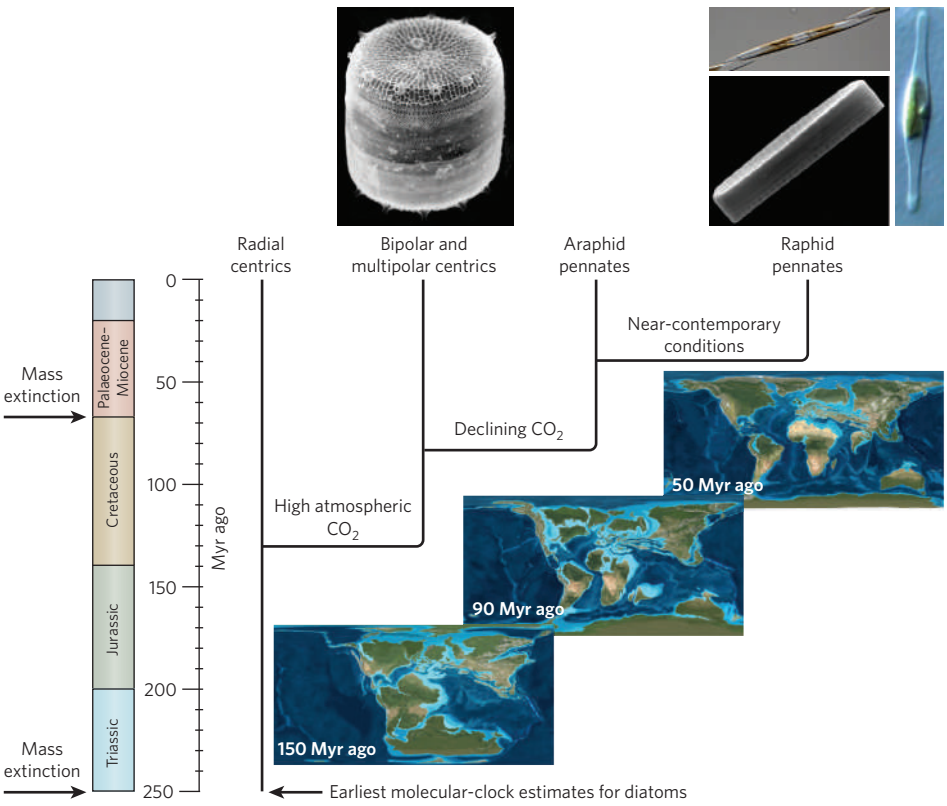


Figure 3 | Estimated timing of divergence of the four major diatom lineages and coincident events in Earth's history. Shown above two of the branches are images of the four species for which the whole genome sequence is available: the multipolar centric *Thalassiosira pseudonana* (courtesy of N. Kroger, Georgia Institute of Technology, Atlanta), and the raphid pennates (from left to right) *Pseudo-nitzschia multiseries* (top; courtesy of K. Holtermann, University of Washington, Seattle), *Fragilariopsis cylindrus* (bottom; courtesy of G. Dieckmann, Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany) and *Phaeodactylum tricornutum* (right; courtesy of C. Bowler, Ecole Normale Supérieure, Paris). To date, neither a representative radial centric nor an araphid pennate has been chosen for whole-genome sequencing. Maps (courtesy of R. C. Blakey, Northern Arizona University, Flagstaff) are palaeographic reconstructions of continent locations during the emergence of the diatom lineages. Shallower depths in the ocean are indicated by lighter blues. Timing of divergence in Myr ago compiled from refs 7 and 26.

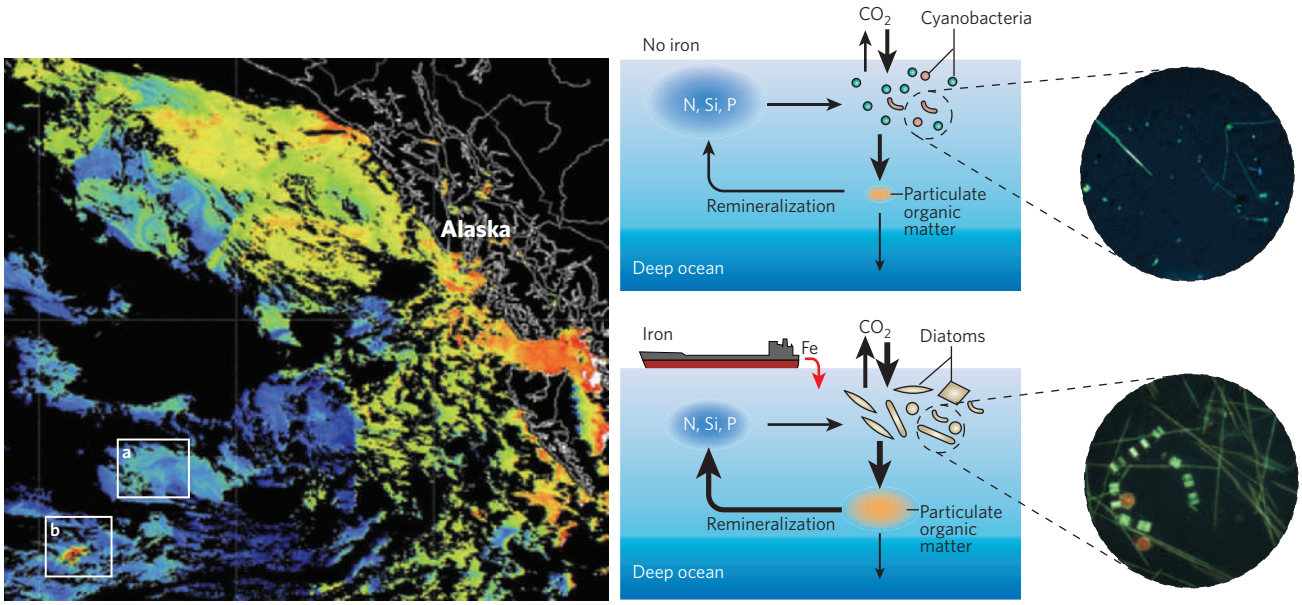


Figure 4 | The effect of iron fertilization on diatoms. The left panel shows a SeaWiFS satellite image of a phytoplankton bloom resulting from iron fertilization in the northeast Pacific Ocean during the SERIES iron-enrichment experiment. The coast of Alaska is shown. Warm colours (reds and yellows) indicate high concentrations of chlorophyll *a* and thus high phytoplankton biomass; cool colours (blues) indicate low chlorophyll *a* concentrations. Dark areas over the ocean result from cloud cover. White boxed regions indicate areas of no iron addition (**a**) and a 700-km^2 region of high chlorophyll *a* concentration, resulting from the addition of iron (**b**). Middle panels show representations of phytoplankton communities and the relative nutrient concentrations present before (top) and after (bottom) the addition of iron to surface waters. The thicker arrows in the bottom panel reflect enhanced carbon fluxes after the addition of iron. The right panels show micrographs of the resultant phytoplankton communities before and after the addition of iron to surface water, collected from a site near that in **b**. In both cases, most of the newly fixed carbon is consumed and respired as CO_2 in the upper ocean. The addition of iron shifts the community from one dominated by small cyanobacteria to one dominated by raphid pennate diatoms such as *Pseudo-nitzschia* (needle-like cells) and centric diatoms (other red or green cells). Both communities were stained with a dye that localizes to newly precipitated silica to illustrate actively dividing diatoms. N, nitrate; P, phosphate; Si, silicic acid. (Left panels courtesy of J. Gower, Orbimage/NASA. Right panels courtesy of C. Durkin and A. Marchetti, University of Washington, Seattle.)

least 30 described coastal and open-ocean species, including some that dominate iron-enrichment experiments in HNLC regions[58]. Only about one-third of known *Pseudo-nitzschia* species and one species of the closely related *Nitzschia* have been shown to produce domoic acid; no other diatoms are known to make toxins. Domoic acid production can be controlled under laboratory conditions, although the absolute concentration produced varies between species, as well as between different strains of the same species; for example, open-ocean strains have not been found to produce significant amounts of domoic acid[59]. Numerous explanations for the observed variation have been proposed[56]. Variation in toxin production among strains could reflect differences in their associated bacterial communities[23], possibly representing another example of interaction between kingdoms[60]. Alternatively, initial analysis of the *Pseudo-nitzschia* genome indicates the presence of a large number of transposable elements, or 'jumping genes'. Movement of these elements to positions near important regulators of domoic acid production could also result in apparently random changes in different strains. The availability of the genome sequence for *P. multiseries* should provide more insights into domoic acid biosynthesis, allowing researchers to profile production capabilities across different species, to better characterize their responses to environmental triggers, and to examine the molecular interactions between toxin-producing cells and bacteria. The genomic data should also help to determine whether toxin production by *Pseudo-nitzschia* is a further example of capabilities being gained directly from another organism.

Seas of change

The oceans are constantly changing. Every day, about 22 million tonnes of atmospheric CO_2 dissolves in the oceans, lowering the pH and changing the chemistry, potentially making it harder for organisms such as bivalves, corals and coccolithophorids to create their calcium carbonate shells[61]. Seasonal variability in the upwelling of deep CO_2-rich waters amplifies the acidification, particularly in the Southern Ocean[62] and coastal upwelling systems[61], where productivity tends to be dominated by diatoms. Ocean waters are also warming[63], wind patterns are shifting, and ocean circulation is changing[64], which together shift turbulent mixing and the delivery of nutrients from deep waters to surface waters[32]. In addition, low-nutrient open-ocean regions seem to be expanding[65]. Rapid warming in the Arctic has thinned and melted sea ice, potentially enhancing phytoplankton productivity as more light penetrates to deeper waters but potentially dampening diatom productivity through changes in the delivery of silicate-rich or nitrate-rich waters to the Arctic. In the Southern Ocean, wind intensity is increasing, affecting the Antarctic Circumpolar Current[64,66]. Wind-driven changes in the speed of this current could shift the delivery of both phytoplankton nutrients and CO_2 from deep waters to surface waters. Globally, populations of apex predators are declining, and nutrient inputs into coastal waters continue to rise[67], giving concern that these changes may be linked to more frequent blooms of toxin-producing phytoplankton[57]. Couple these human-induced changes with the effect of natural climatic oscillations[68] and the urgent need for better monitoring of marine ecosystems becomes clear[69,70].

How will critical components of marine food webs, such as diatoms, respond to such large changes occurring over a relatively short time? Some predict that diatoms will have a greatly reduced role in future phytoplankton communities. In this scenario, the ocean would be dominated by cyanobacteria, green algae[32] and coccolithophorids[71], groups well adapted to compete in the low-nutrient environments characteristic of a less turbulent ocean. A shift away from diatom-based communities would bring a dramatic reduction in the ability of ocean biota to sequester CO_2 from the atmosphere, exacerbating climate change[71]. However, diatoms are masters at surviving in a wide variety of conditions, including in highly stratified, nutrient-poor regions such as the North Pacific Gyre[72]. Even so, there will be changes to the distribution

of some species of diatom and perhaps to the timing of diatom blooms. Many species will adapt to the changing ocean environment, but others will decline in abundance and some will disappear.

The future of diatom research

It is important to know how diatoms affect ocean ecology and biogeochemistry at any given time in any given region. Sequencing the genomes of additional representative diatoms, in combination with analysing the genomes of diatom communities in nature, will identify the core attributes that allowed these organisms to cope with past conditions and will help to interpret responses to today's conditions. Next-generation ecogenomic sensors (see page 180), which continuously monitor the presence of sentinel species or the expression of sentinel genes, are needed to provide information about global patterns of biologically relevant physiochemical properties. Continuously monitoring the genes encoding the iron-storage molecule ferritin, for example, would provide information about the presence and the biological availability of iron in surface waters, which both seem to be changing[73]. This increasingly genomic approach will make it possible to move beyond speculation about the state of the environment to instead document the changes actually occurring in critical groups such as diatoms before they become the new canaries in the coal mine. ■

1. Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biogeochem. Cycles* **9,** 359–372 (1995).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281,** 237–240 (1998).
3. Sarthou, G., Timmermans, K. R., Blain, S. & Treguer, P. Growth physiology and fate of diatoms in the ocean: a review. *J. Sea Res.* **53,** 25–42 (2005).
4. Denman, K. L. Climate change, ocean processes and ocean iron fertilization. *Mar. Ecol. Prog. Ser.* **364,** 219–225 (2008).
5. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306,** 79–86 (2004).
This paper describes results from the first diatom genome-sequencing project.
6. Bowler, C. *et al.* The *Phaeodactylum* reveals the evolutionary history of diatom genomes. *Nature* **456,** 239–244 (2008).
This paper provides the first whole-genome comparison of two diatoms and identifies the presence in the genome of numerous bacterial genes.
7. Kooistra, W. H. C. F., Gersonde, R., Medlin, L. K. & Mann, D. G. in *Evolution of Primary Producers in the Sea* (eds Falkowski, P. G. & Knoll, A. H.) 207–249 (Elsevier, 2007).
8. Morel, F. M. M. & Price, N. M. The biogeochemical cycles of trace elements in the oceans. *Science* **300,** 944–947 (2003).
9. Villareal, T. A. *et al.* Upward transport of oceanic nitrate by migrating diatom mats. *Nature* **397,** 423–425 (1999).
10. Vardi, A. *et al.* A diatom gene regulating nitric-oxide signalling and susceptibility to diatom-derived aldehydes. *Curr. Biol.* **18,** 895–899 (2008).
11. Ianora, A. *et al.* Aldehyde suppression of copepod recruitment in blooms of a ubiquitous planktonic diatom. *Nature* **429,** 403–407 (2004).
12. Parker, M. S., Mock, T. & Armbrust, E. V. Genomic insights into marine microalgae. *Annu. Rev. Genet.* **42,** 619–645 (2008).
13. Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattachrya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21,** 809–818 (2004).
14. Reyes-Prieto, A., Hackett, J. D., Soares, M. B., Bonaldo, M. F. & Bhattacharya, D. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr. Biol.* **16,** 2320–2325 (2006).
15. Huang, J. & Gogarten, J. P. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* **8,** R99 (2007).
16. Becker, B., Hoef-Emden, K. & Melkonian, M. Chlamydial genes shed light on the evolution of phototrophic eukaryotes. *BMC Evol. Biol.* **8,** 203 (2008).
17. Allen, A. E., Vardi, A. & Bowler, C. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* **9,** 264–273 (2006).
18. Montsant, A. *et al.* Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J. Phycol.* **43,** 585–604 (2007).
19. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. *Nature* **438,** 90–93 (2005).
20. Droop, M. R. Vitamins, phytoplankton and bacteria: symbiosis or scavenging. *J. Plankton Res.* **29,** 107–113 (2007).
21. Reimann, L., Steward, G. F. & Azam, F. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl. Environ. Microbiol.* **66,** 578–587 (2000).
22. Lau, W. W. Y., Keil, R. G. & Armbrust, E. V. Succession and diel transcriptional response of the glycolate-utilizing component of the bacterial community during a spring phytoplankton bloom. *Appl. Environ. Microbiol.* **73,** 2440–2450 (2007).
23. Kaczmarska, I. *et al.* Diversity and distribution of epibiotic bacteria on *Pseudo-nitzschia multiseries* (Bacillariophyceae) in culture, and comparison with those on diatoms in native seawater. *Harmful Algae* **4,** 725–741 (2005).
24. Foster, R. A. & Zehr, J. P. Characterization of diatom–cyanobacteria symbioses on the basis of *nifH*, *hetR* and 16S rRNA sequences. *Environ. Microbiol.* **8,** 1913–1925 (2006).
25. Schmid, A.-M. M. Endobacteria in the diatom *Pinnularia* (Bacillariophyceae). I. 'Scattered ct-nucleoids' explained: DAPI–DNA complexes stem from exoplastidial bacteria boring into the chloroplasts. *J. Phycol.* **39**, 122–138 (2003).
26. Sorhannus, U. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Mar. Micropaleontol.* **65,** 1–12 (2007).
27. Sims, P. A., Mann, D. G. & Medlin, L. K. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* **45,** 361–402 (2006).
This paper provides detailed information that couples geological data with diatom morphology and palaeological distributions.
28. Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305,** 354–360 (2004).
29. Katz, M. E. *et al.* Biological overprint of the geological carbon cycle. *Mar. Geol.* **217,** 323–338 (2005).
30. Falkowski, P. G. *et al.* The rise of oxygen over the past 205 million years and the evolution of large placental mammals. *Science* **309,** 2202–2204 (2005).
31. Guidry, M. W., Arvidson, R. S. & MacKenzie, F. T. in *Evolution of Primary Producers in the Sea* (eds Falkowski, P. G. & Knoll, A. H.) 377–403 (Elsevier, 2007).
32. Falkowski, P. G. & Oliver, M. J. Mix and max: how climate selects phytoplankton. *Nature Rev. Microbiol.* **5,** 813–819 (2007).
33. Rabosky, D. L. & Sorhannus, U. Diversity dynamics of marine planktonic diatoms across the Cenozoic. *Nature* **457,** 183–187 (2009).
34. Zielinski, U. & Gersonde, R. Diatom distribution in Southern Ocean surface sediments (Atlantic sector): implications for paloeenvironmental reconstructions. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **129,** 213–250 (1997).
35. Kroth, P. G. *et al.* A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3,** e1426 (2008).
36. Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Iron cycling and nutrient-limitation patterns in surface waters of the world ocean. *Deep-Sea Res. II* **49,** 463–507 (2002).
37. Marchetti, A., Maldonado, M. T., Lane, E. S. & Harrison, P. J. Iron requirements of the pennate diatom *Pseudo-nitzschia*: comparison of oceanic (HNLC) and coastal species. *Limnol. Oceanogr.* **51,** 2092–2101 (2006).
38. Sunda, W. G., Swift, D. G. & Huntsman, S. A. Low iron requirement for growth in oceanic phytoplankton. *Nature* **351,** 55–57 (1991).
39. Strzepek, R. F. & Harrison, P. J. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431,** 689–692 (2004).
This paper provides the first molecular-based description of photosynthetic differences between coastal and open-ocean diatoms.
40. Peers, G. & Price, N. M. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441,** 341–344 (2006).
41. Kustka, A. B., Allen, A. E. & Morel, F. M. M. Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms. *J. Phycol.* **43,** 715–729 (2007).
42. Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105,** 10438–10443 (2008).
43. Jickells, T. D. *et al.* Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science* **308,** 67–71 (2005).
44. Marchetti, A. *et al.* Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature* **457,** 467–470 (2008).
This was the first molecular description of an iron-storage protein in diatoms and discusses the potential selective advantages conveyed by this protein.
45. Drum, R. W. & Gordon, R. Star Trek replicators and diatom nanotechnology. *Trends Biotechnol.* **21,** 325–328 (2003).
46. Tréguer, P. *et al.* The silica balance in the world ocean: a reestimate. *Science* **268,** 375–379 (1995).
47. Vrieling, E. G. *et al.* Salinity-dependent diatom biosilicification implies an important role of external ionic strength. *Proc. Natl Acad. Sci. USA* **104,** 10441–10446 (2007).
48. Bidle, K. D., Maganelli, M. & Azam, F. Regulation of oceanic silicon and carbon preservation by temperature control on bacteria. *Science* **298,** 1980–1984 (2002).
49. Kroger, N., Lorenz, S., Brunner, E. & Sumper, M. Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science* **298,** 584–586 (2002).
This paper describes the isolation of native silaffins from the cell wall and the function of these proteins during silica precipitation.
50. Kroger, N., Deutzmann, R., Bergsdorf, C. & Sumper, M. Species-specific polyamines from diatoms control silica morphology. *Proc. Natl Acad. Sci. USA* **97,** 14133–14138 (2000).
51. Wenzel, S., Hett, R., Richthamer, P. & Sumper, M. Silacidins: highly acidic phosphopeptides from diatom shells assist in silica precipitation *in vitro*. *Angew. Chem. Int. Edn Engl.* **120,** 1753–1756 (2008).
52. Mock, T. *et al.* Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioproceses. *Proc. Natl Acad. Sci. USA* **105,** 1579–1584 (2008).
53. Siegenthaler, U. *et al.* Stable carbon cycle–climate relationship during the late Pleistocene. *Science* **310,** 1313–1317 (2005).
54. Matsumoto, K. & Sarmiento, J. L. A corollary to the silicic acid leakage hypothesis. *Paleoceanography* **23,** PA2203 (2008).
55. Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science* **315,** 612–617 (2007).
56. Bates, S. S. & Trainer, V. L. in *Ecology of Harmful Algae* (eds Graneli, E. & Turner, J. T.) Ch. 7 (Springer, 2006).
57. Goldstein, T. *et al.* Novel symptomatology and changing epidemiology of domoic acid toxicosis in California sea lions (*Zalophus californianus*): an increasing risk to marine mammal health. *Proc. R. Soc. B* **275,** 267–276 (2008).
58. de Baar, H. J. W. *et al.* Synthesis of iron fertilization experiments: from the iron age in the age of enlightenment. *J. Geophys. Res.* **110,** C09S16 (2005).
59. Marchetti, A. *et al.* Identification and assessment of domoic acid production in oceanic *Pseudo-nitzschia* (Bacillariophyceae) from iron-limited waters in the northeast subarctic Pacific. *J. Phycol.* **44,** 650–661 (2008).
60. McDougald, D., Rice, S. A. & Kjelleberg, S. Bacterial quorum sensing and interference by

naturally occurring biomimics. *Anal. Bioanal. Chem.* **387,** 445–453 (2007).
61. Feely, R. A., Sabine, C. L., Hernandez-Ayon, J. M., Ianson, D. & Hales, B. Evidence for upwelling of corrosive 'acidified' water onto the continental shelf. *Science* **320,** 1490–1492 (2008).
62. McNeil, B. I. & Matear, R. J. Southern Ocean acidification: a tipping point at 450-ppm atmospheric CO_2. *Proc. Natl Acad. Sci. USA* **105,** 18860–18864 (2008).
63. Levitus, S., Antonov, J. I., Boyer, T. P. & Stephens, C. Warming of the world ocean. *Science* **287,** 2225–2229 (2000).
64. Toggweiler, J. R. & Russell, J. Ocean circulation in a warming climate. *Nature* **451,** 286–288 (2008).
65. Polovina, J. J., Howell, E. A. & Abecassis, M. Ocean's least productive waters are expanding. *Geophys. Res. Lett.* **35,** L03618 (2008).
66. Boning, C. W., Dispert, A., Visbeck, M., Rintoul, S. R. & Schwarzkopf, F. U. The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geosci.* **1,** 864–869 (2008).
67. Jackson, J. B. C. Ecological extinction and evolution in the brave new ocean. *Proc. Natl Acad. Sci. USA* **105,** 11458–11465 (2008).
68. Cloern, J. E., Jassby, A. D., Thompson, J. K. & Hieb, K. A. A cold phase of the East Pacific triggers new phytoplankton blooms in San Francisco Bay. *Proc. Natl Acad. Sci. USA* **104,** 18561–18565 (2007).
69. Richardson, A. J. & Poloczanska, E. S. Under-resourced, under threat. *Science* **320,** 1294–1295 (2008).
70. Smetacek, V. & Cloern, J. E. On phytoplankton trends. *Science* **319,** 1346–1348 (2008).
71. Cermeno, P. *et al.* The role of nutricline depth in regulating the ocean carbon cycle. *Proc. Natl Acad. Sci. USA* **205,** 20344–20349 (2008).
This paper predicts phytoplankton community structure in oceans of the future and the resultant effects on the carbon cycle.
72. Dore, J. E., Letelier, R. M., Church, M. J., Lukas, R. & Karl, D. M. Summer phytoplankton blooms in the oligotrophic North Pacific Subtropical Gyre: historical perspective and recent observations. *Prog. Oceanogr.* **76,** 2–38 (2008).
73. Sedwick, P. N., Sholkovitz, E. R. & Church, T. M. Impact of anthropogenic combustion emissions on the fractional solubility of aerosol iron: evidence from the Sargasso Sea. *Geochem. Geophys. Geosyst.* **8,** Q10Q06 (2007).

Acknowledgements I am grateful to members of my laboratory and to G. Rocap and S. Francis for discussion and edits of the manuscript. Support was provided by funding from the Gordon and Betty Moore Foundation Marine Microbiology Initiative, the US National Science Foundation and the National Institute of Environmental Health Sciences.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (armbrust@ocean.washington.edu).

Microbial community structure and its functional implications

Jed A. Fuhrman[1]

Marine microbial communities are engines of globally important processes, such as the marine carbon, nitrogen and sulphur cycles. Recent data on the structures of these communities show that they adhere to universal biological rules. Co-occurrence patterns can help define species identities, and systems-biology tools are revealing networks of interacting microorganisms. Some microbial systems are found to change predictably, helping us to anticipate how microbial communities and their activities will shift in a changing world.

Microorganisms, by which I mean Bacteria, Archaea, viruses, protists and fungi, are vital to the function of all ecosystems. This is largely because they exist in enormous numbers (there are roughly 5×10^{30} bacteria alone worldwide)[1] and so have immense cumulative mass and activity[1]. They are also probably more diverse than any other organisms, so it is easy to see why the structure of microbial communities, that is, the different kinds of organisms and their abundances, is so important to the way in which ecosystems function. But even with modern tools, it is not easy to determine microbial community structure and map its variations in space and time; we have only recently begun to learn about the scales of variation (Box 1). Understanding ecosystem function, and predicting Earth's response to global changes such as warming and ocean acidification, calls for much better knowledge than we have today about microbial processes and interactions.

In the past couple of decades, the use of genome sequences and related approaches[2–6] has overcome the need for cultivation to characterize and identify microorganisms in nature (Box 2). It still seems almost hopeless to sort out the identities and interrelationships among the trillions of microorganisms in a cubic metre of sea water, let alone a few hectares of ocean, but high-throughput sequencing and whole-community fingerprinting techniques have enabled researchers to make considerable progress (Box 2). This review summarizes our knowledge of microbial community structure, with a focus on planktonic marine bacteria, and discusses what we can learn about microbial systems and their functions from this information (Box 3).

General distributions and the 'rare biosphere'

The classic dictum about microbial distribution patterns, "everything is everywhere but the environment selects", is attributed to Lourens Baas Becking[7]. This concept has been thoughtfully reviewed in recent years[8–12], often with particular attention to the long tail of the species abundance curve, which shows that large numbers of individually rare species are found in most ecosystems (Fig. 1). This concept reflects the fact that current distributions of organisms are the result of historical factors, including dispersion by wind, water and animals, and adaptations to local conditions that change over space and time. The 'everything is everywhere' part alludes to the remarkable dispersal potential of microorganisms. It has been claimed, largely on the basis of studies of some morphologically defined protistan species, that for species comprising organisms that are smaller than 1 mm, global diversity is relatively low and organisms are cosmopolitan, essentially lacking biogeographical variations[13]. This is presumably owing to their high population sizes and easy dispersion. It is impossible to disprove this assertion, because we cannot prove the complete absence of an organism. Nevertheless, in practice and to the extent that most measurements allow (Box 2), numerous studies suggest that most microorganisms do not seem to be cosmopolitan, even within a given habitat type, and discernible biogeographical patterns are typical[9,10,14]. The idea that 'the environment selects' indicates that only those organisms capable of activity and growth in a particular environment will increase in number. There is a fuzzy boundary between 'common' and 'rare' organisms, often described as somewhere in the range of 0.1–1% of the microbial community. It has been argued that the rare ones do not ordinarily affect most major biogeochemical processes, such as respiration and the processing of nutrients[8]. Studies indicate that in marine plankton, common organisms typically carry out most of the activity[15,16]. Cultivation techniques often yield bacteria that are very rare in whole-community cloning studies, such as *Vibrio* spp. By contrast, most molecular studies tend to find primarily the common organisms[8], and characterizing the rare ones with molecular tools typically requires a focused study, such as high-throughput tag sequencing[17,18].

Among the more common, and presumably more active, microorganisms, molecular survey data can be used to show the extent to which particular organisms are cosmopolitan, widespread or endemic within a given habitat type and can relate distributions to other properties. One study[19] of marine-plankton clone libraries collected samples from nine widespread locations, with 263–702 clones each, and coverage estimates in the range 45–94%. Of 582 unique operational taxonomic units (OTUs), defined in that study as having at least 97% 16S ribosomal RNA similarity, 69% were found only at a single location (endemic), 17% were at two locations, 6% at three locations, and only 0.4% were cosmopolitan (found at all nine locations). The proportion of endemic ones was similar, irrespective of community size. The more widespread OTUs also tended to be the most abundant at individual locations. The endemic OTUs tended to be individually rare; 92% of them had only 1–2 clones, but four endemic OTUs were relatively abundant, with 7–24 clones at their respective locations. A simplistic model of these data is that an organism abundant in one place is likely to be detectable and may be abundant elsewhere, whereas one that is rare but detectable (0.1–1% by the approach used in this study) in one location is likely to be below the detection limit in other locations. Many of these rare ones are presumably present elsewhere but are too rare to detect easily, or they may become abundant at some other place or time. The apparent relationship between abundance and range size is one of several

[1]Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA.

instances in which microbial patterns parallel those of larger organisms[14] (discussed further in the section 'Universal patterns').

The apparently low contribution of rare organisms to most biogeochemical processes does not mean they are unimportant. Yes, in a given environment, most may simply be passing through and may die without ever finding a suitable niche. But sometimes the highly diverse members of the 'rare biosphere'[17] can have clear ecological significance. The most obvious influence is by acting like a seed bank[8]. Organisms that may be ideally adapted to conditions in another time or place could eventually thrive by just dispersing and waiting. Many microorganisms that are rare in a given plankton sample might be in the process of dispersion among adjacent niches in which they may be extremely abundant; for example, in suspended aggregates of 'marine snow', in the guts or on the surfaces of animals (for instance, pathogens such as *Vibrio vulnificus*), or in sediments. Many that are rare at a given time probably thrive during a different season or in occasional phenomena such as El Niño events. Marine bacteria that are rare in one season can be abundant in another. For example, in a four-year time-series study[20], a variety of taxa were undetectable in some months (<0.1% of the total), but then made up several per cent of the community in other months. In some cases, the seed-bank aspect may apply not to whole organisms but just to their genes[17]. A rare individual may have some genes that when transferred to another organism will create a recombinant that is better adapted to a particular habitat than either parent. The seed bank can function as a valuable insurance as global and local conditions change through natural or anthropogenic causes.

Even some 'chronically rare' organisms may have global biogeochemical significance. There is no mechanism to prevent the long-term persistence of widespread but rare organisms, and it stands to reason that rare organisms are probably growing slowly (otherwise they would become common). It has been argued that being rare and slow-growing helps them avoid major mortality processes such as viral infection and predation[8]. The viral infection rate is directly dependent on the abundance of hosts. Predation by protists generally removes the larger organisms that tend to be growing faster, so rare ones are presumably smaller and grazed on less[8]. Chance extinction as a result of stochastic events is reduced by enormous microbial population sizes; planktonic bacteria, representing only 1 in 10,000 cells, still have 10^{14} individuals per cubic kilometre. In addition, rare organisms can collectively affect globally important biogeochemical processes. For example, marine nitrogen fixation — a significant input of nitrogen into the global biosphere — was thought to be performed primarily by *Trichodesmium*, a globally rare but locally (and episodically) abundant colonial cyanobacterium[21], or by symbiotic cyanobacteria in larger phytoplankton[22]. But recent evidence suggests that chronically rare cells, which represent about 1 in 10,000 of the total marine bacteria, may at times cumulatively fix more nitrogen than the larger organisms[23]. Perhaps other rare organisms with unique capabilities are important for processes such as slowly breaking down recalcitrant dissolved organic compounds in sea water.

Box 1 | Scales of variation

Changes in the community structure in space and time are very informative. For a start, they show us what scales a particular sample represents. This is crucial for extrapolating from individual samples to the world at large, a common exercise that is often not well informed. For example, samples collected for the *Sorcerer II* Global Ocean Sampling (GOS) Expedition[74] were typically tens of litres of sea water from near the surface. What region of the ocean (horizontal and vertical), and over what timescale, might the microorganisms in such a sample represent quantitatively? Recent community fingerprinting studies with similar samples of marine plankton have found timescales for significant community changes to be typically of the order of days and weeks[51,75]. In terms of size scales, patches of coherent communities at a given depth horizon tend to be on the scale of kilometres or tens of kilometres[51,75]. In the vertical direction, there can be significant changes over metres or tens of metres, or even over millimetres at the immediate sea-surface microlayer[76]. This suggests a given GOS sample is likely to represent a range of about a week temporally, a few kilometres horizontally and a few metres or tens of metres vertically. But beyond this practical application, changes in community structure help us to understand factors that control communities. For example, the horizontal 10-km scale of variability resembles that of mesoscale ocean eddies of factors such as chlorophyll, as observed by satellite[51], suggesting that the physicochemical and biological factors that structure the microbial community composition may relate to, or even be the same as, those that control phytoplankton abundance.

Difficulties evaluating dispersion

The distribution of organisms is controlled by dispersion and adaptations to environmental heterogeneity[14], so understanding dispersion is important. We can demonstrate that dispersion has happened by finding the same kind of organism in different places. But the OTUs used to identify 'sameness' in microbial studies are not equivalent to species in larger organisms. Frederick Cohan[24] suggested conservatively that what is called a single species in bacteria might be as broad as a genus in the macrobiota, and James Staley[25] made the point that the formal definition of a bacterial species would classify all the primates from humans to lemurs as a single species. There is not a precise 16S rRNA 'molecular clock', but a 1% change in bacterial 16S sequence is estimated to take millions of years[26,27]. This is important in evaluating the role of dispersion, because what we call the 'same' microorganism, on the basis of membership in an OTU or even an accepted species, may be so varied as to be irrelevant for dispersion studies. So we should probably be considering 'microdiversity'[28–30] when looking for dispersion as a mechanism to explain contemporary distributions. One would expect that only individuals that are essentially identical in slowly evolving phylogenetic markers such as 16S rRNA would be suitable as evidence for recent dispersion.

Universal patterns

Do microorganisms follow well-established patterns previously observed in larger organisms, implying the presence of 'universal' processes or profound unifying rules that apply to all life? Or is there something fundamentally different in character about these smallest organisms that allows us to distinguish between rules that are truly universal and those that are not? Growing evidence suggests that microorganisms do follow some classic ecological patterns, bolstering the likelihood of successfully applying established ecological theory[31]. Examples are latitudinal gradients in diversity, taxa–area relationships and community assembly 'rules'.

Latitudinal gradients

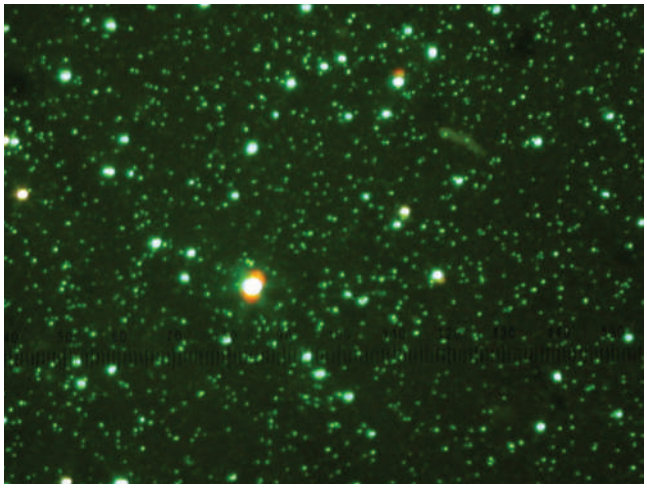
One of the oldest observed patterns of animal and plant diversity, first reported in the early nineteenth century, is the tendency of lower latitudes to have more species than higher latitudes[32–34]. There are several possible causes that are not mutually exclusive and continue to be debated[32–34]. One hypothesis is that on land, generally higher productivity in lower latitudes provides more resources that can be split into more niches. Another hypothesis is that higher temperatures in low latitudes increase the metabolic rate and make biological processes, including speciation, occur faster. Does this latitudinal gradient apply to microorganisms, and if so, can the particular pattern help us choose between the hypotheses? Some reports suggest that there is little or no microbial gradient[13]. Meta-analysis of data on a wide variety of species (from protists to megafauna)[34] or exclusively on marine species[35] indicated that the strength and slope of the latitudinal richness gradient is reduced as organism size decreases, which could be extrapolated to suggest that there is little or no gradient for bacteria-sized organisms. A fingerprinting study of soil bacteria found a moderate range of richness, best predicted by soil pH, with no discernible latitudinal gradient[36]; acidic soils had much lower diversity than neutral or alkaline soils, irrespective of latitude and temperature. In this case, it seems that an overriding influence of soil pH combined with high natural variability in soil habitats (and differing sampling effort[14]) could have masked other, more subtle, influences on diversity.

In contrast to these results, two recent reports found a bacterial latitudinal richness gradient in marine plankton. One was a study of

Box 2 | Measuring and comparing community composition

Most contemporary studies of microbial community composition are DNA-based, avoiding cultivation, which misses most organisms[77]. DNA extracted from the microbial community (the 'metagenome') can be analysed in several ways:
(1) by cloning and sequencing phylogenetically informative genes, such as those for 16S ribosomal RNA;
(2) by analysing all the genes studied, either by random cloning into large or small insert libraries and sequencing those, or by clone-free methods such as pyrosequencing;
(3) by using high-throughput sequence analysis of phylogenetically informative short sequence tags[17];
(4) by using whole-community fingerprinting methods such as terminal restriction fragment-length polymorphism (TRFLP), denaturing gradient gel electrophoresis (DGGE) or automated ribosomal intergenic spacer analysis (ARISA). These less costly fingerprinting methods yield data (such as gene fragment sizes for TRFLP and ARISA) on specific components of the community, but provide less information about individuals and usually allow the putative identification of components by comparison with sequences. The organisms from which the gene sequences or fingerprints were derived are categorized, or 'binned', into operational taxonomic units (OTUs) on the basis of a variety of criteria that can vary with the method, such as sequence similarity or fragment size. Some studies do not use extracted DNA but instead involve tagging and counting particular organisms (one OTU at a time) by fluorescence *in situ* hybridization. With all these methods, quantitative interpretation of the results must consider possible biases, which can occur at several steps from collection to binning[77]. Some approaches are reasonably quantitative for certain OTUs, such as abundance estimates of *Prochlorococcus* by ARISA[20]. Even imperfect methods, like some used in ecology (for example, fogging a tree with insecticide to get 'all' the insects to fall into a net), have been used to develop fundamental theories.

Comparing community composition between samples is challenging when the assay covers only a minority of each community, as occurs in highly diverse samples analysed by limited numbers of sequences. High sequence coverage helps and, for moderately diverse samples such as marine planktonic bacteria, a few hundred 16S rRNA clones are estimated to have about 50–90% OTU coverage[19]. Alternatively, whole-community fingerprinting, by methods such as ARISA, which can show essentially all the bacterial OTUs making up more than 0.1% of the community[72], allows cost-efficient comparison of multiple samples[36,72,75], as long as one can ignore organisms that each represent less than 0.1% of the population. Similarly, comparing OTU richness (the total number of measured OTUs) or species lists between samples is challenging because each protocol catches a different proportion of the rare OTUs (no protocol can catch them all). They therefore typically compare lists with different levels of completeness, even with a massive sequencing effort[18]. The fairest comparisons use standardized measures of composition or richness, such as by fingerprinting[37], or a uniform and high-coverage technique to generate sequences. This problem is well known in ecology, and statistical approaches have been used[17,72] and continue to be developed[73,78,79] to estimate richness and compare diversity while taking into account the varying amounts of undetected organisms.



The image is an epifluorescence micrograph of marine viruses (smallest green dots), bacteria and/or archaeal organisms (medium-size green dots) and pigmented protists (larger green dots with red patches), representing a small part of a microlitre of sea water from the northeast Pacific Ocean off California. For scale, the bacteria average about 0.5 μm in diameter. The non-descript nature of these organisms means that we need molecular techniques to characterize the community composition.

nine samples using 16S rRNA clone libraries[19], and the other was a whole-community fingerprinting study[37] that examined 103 samples from near the sea surface at 56 locations worldwide. In the latter study, the measured richness values, when plotted against latitude, fell within triangular constraint envelopes whereby high latitudes had low richness, but lower latitudes could have high or low richness. This suggests that latitude sets an upper limit on richness but that other factors can sometimes reduce it. In both studies, richness was correlated with temperature, but not with indices of productivity such as chlorophyll or annual primary production, consistent with the theory that the metabolic rate, affected by temperature, may strongly influence the pace of complex ecological processes such as speciation[37]. Why is this different from soils? Perhaps by studying near-surface plankton in sea water, a habitat much more uniform than soils, many chemical and physical factors were held relatively constant, allowing the detection of an effect related to latitude despite the presence of other important factors.

Taxa–area relationships

Another widely studied ecological pattern is the relationship between the number of observed species (S) and the sampled area (A), commonly assumed to have a power-law relationship, $S \propto A^z$. Values for z are determined empirically, and for animals and plants are typically 0.1–0.3 for contiguous habitats and 0.25–0.35 for islands[33]. Studies of microorganisms (including marine planktonic diatoms and salt-marsh bacteria, but not planktonic bacteria) have found a range of z values, often lower than 0.1, but some recent studies have reported z values that overlap the canonical ranges, especially for 'island-like' situations[12,14,31,38,39]. The broader definition of a microbial (as opposed to animal or plant) species lowers z significantly[38], and z is also affected by the difficulty in comprehensively sampling microorganisms[14,39]. Nevertheless, the overlap in values suggests that relationships for microorganisms can be similar quantitatively to those of larger organisms, indicating that this may be a universal rule for all domains of life.

Related to taxa–area relationships are distance–decay relationships, which show community differences with increasing distance[14]. For deep-sea bacteria, fingerprinting shows a significant and steady decline in community similarity over a 3,500-km distance at a depth of 3,000 m (in the Pacific Ocean), and over a 1,000-km distance at a depth of 1,000 m (Atlantic Ocean near the Amazon plume), with much scatter at 1,000 m (Pacific) and 500 m (Atlantic)[40]. These results dispel the notion that the deep sea is uniform in microbial communities, and show that different depths and locations have different relationships, with considerable patchiness. One hypothesis[40] is that differences in patterns may relate to shallower and mid-water depths being most influenced by the patchy 'raining' of detritus that provides food to the deep sea, with most organic carbon being consumed before it reaches the lower depths.

Community assembly 'rules'

Co-occurrence patterns of organisms — examining which organisms sometimes or never occur together — have been used to reveal community assembly rules[41,42]. Several ecological processes potentially contribute to these nonrandom co-occurrence patterns, including competition,

Box 3 | Applications of microbial community-structure studies

Microbial community structure can be used to address several timely issues.

- Finding basic patterns in microbial distributions, such as the extent of endemism and ubiquity, and learning how to evaluate these given the enormous 'rare biosphere' and difficulties characterizing dispersion
- Searching for general patterns that might apply to all life and support the concept of 'universal' rules that broadly control biology, thereby improving our theories and leading to predictions
- Using co-occurrence patterns to help define potentially interacting organisms and interaction networks in highly complex systems
- Using co-occurrence of organisms and correlations to environmental parameters to help delineate microbial species
- Examining seasonally or annually repeating patterns, also leading to predictions
- Evaluating the roles of viruses
- Linking community structure or particular genes to particular functions

habitat filtering, historical effects and neutral processes[43]. One study[44] compared co-occurrence in more than 100 microbial (Bacteria, Archaea, protists and fungi) data sets with a meta-analysis of almost 100 macroorganism data sets. The microbial assemblages had nonrandom patterns of co-occurrence broadly similar to those in macroorganisms. The authors concluded that some co-occurrence patterns may be general characteristics of all domains of life. The extent of co-occurrence in microbial communities did not vary between broad taxonomic groups or habitat types. There were variations from using different methods to survey microbial communities (such as clone libraries or fingerprinting), and taxonomic resolution was also a factor. The authors also noted that undersampling of microorganisms may underestimate the extent of community structure, so microorganisms might have more highly structured communities than macroorganisms.

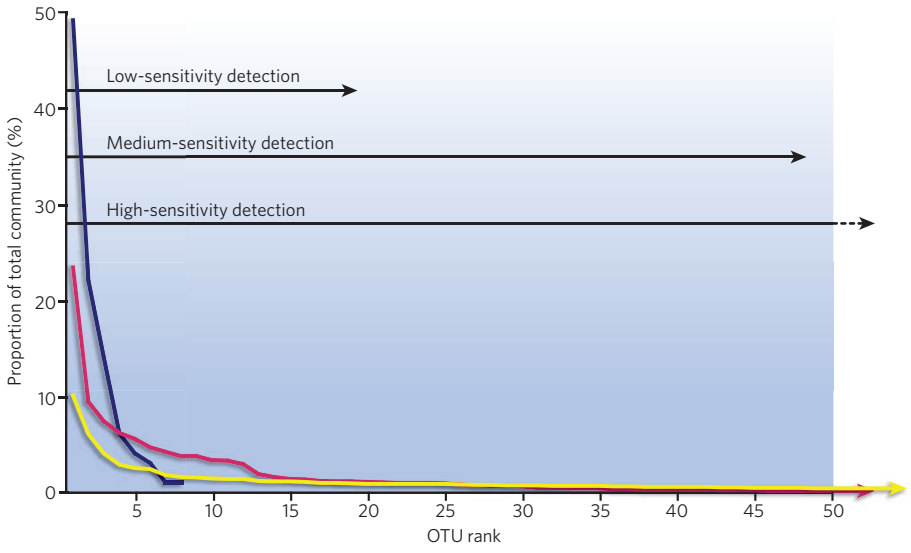


Figure 1 | Rank abundance relationships for bacterial operational taxonomic units. Different communities vary immensely in species richness. The examples here represent an extremely acidic habitat with low diversity (blue line)[71], marine plankton with moderate diversity (red line)[72] and sediment with high diversity (yellow line)[73]. All environments are dominated by just a few operational taxonomic units (OTUs), but have a long tail in the distribution (rare taxa) that can number in the tens of thousands, off the figure to the right. Low-sensitivity methods such as 100-clone libraries show only the most abundant taxa, but that may be sufficient for extremely low-diversity samples. Methods with moderate sensitivity such as 1,000-clone libraries or ARISA (automated ribosomal intergenic spacer analysis) fingerprinting show a larger fraction of the community, but still can miss many rare taxa. High-throughput methods such as tag sequencing[17] can detect thousands of taxa. It is a practical impossibility to detect all of the taxa. Different studies require knowledge of more or less of these rare taxa to support their conclusions, so it is important to select the appropriate method.

Interactions and networks

Microbial communities, as part of natural ecosystems, are inherently complex. The traditional tools of microbiology, such as pure cultures and genetic studies, tend to provide a reductionist view, studying each organism in isolation. However, a reductionist approach is not well suited for learning about interactions and emergent properties of communities. Holistic approaches, which study natural habitats directly, can yield complementary data to help deduce the interactions. Co-occurrence patterns, as described above, also show how particular organisms in a system occur together and vary with environmental parameters. These patterns show important details of a particular community structure and can be represented as mathematical interaction diagrams or networks. If environmental conditions are included in the co-occurrence patterns, the results indicate which conditions the co-occurring assemblages of organisms prefer or avoid. The inclusion of biogeochemical rate measurements may link groups of organisms with particular functions. These kinds of result allow us to examine the potential interactions between organisms and aspects of the niches of microorganisms within extremely complex and dynamic natural communities.

Such microbial network studies are at an early stage and, given the high natural variability, they require a large amount of data if the findings are to be statistically significant. They also require suitable analytical tools for the evaluation and sorting of enormous numbers of potential interactions. Some of the initial studies[45,46] used data from the San Pedro Ocean Time-Series (SPOT) microbial observatory site (southern California). Quansong Ruan and colleagues[45] developed 'local similarity analysis' to evaluate contemporaneous and time-lagged correlations between parameters, and discerned a portion of a network using this approach. Time-lagged relationships (for example between predators and prey) are common in ecology. Joshua Steele and I[46] extended this analysis using tools from systems biology to visualize the results as networks centred on members of particular groups of organisms connected to their 'nearest neighbours', which are organisms or parameters that directly correlate positively or negatively. An example of such a network, centred on members of the ubiquitous SAR11 cluster, shows ten different members of this cluster, each with distinctive combinations of positive and negative correlations to other bacteria or environmental parameters (Fig. 2). Previous work had shown the seasonality of a few broadly defined SAR11 types[47], but this network analysis provides much more detailed distinctions. Interestingly, only two pairs of SAR11 subtypes correlated to each other, and the higher correlated pair were members of different subclades. This demonstrates that ecological relatedness may not follow phylogenetic relatedness, even in a narrow phylogenetic group.

In these networks, the 'interactions' are mathematical relationships (correlations or anticorrelations) that need further investigation to distinguish direct interactions from indirect ones. This is an important distinction, partly because classically studied ecological networks, on which there is an extensive body of work[48], consider only direct interactions. Positive correlations in these mathematically derived networks could be common preferred conditions or perhaps cooperative activities such as cross-feeding. Similarly, negative correlations may represent opposite seasonality, competition for limited resources or perhaps active negative interactions such as targeted allelopathy[49] or predator–prey relationships[50]. The ability of local similarity analysis to see lagged relationships, showing when one parameter changes in

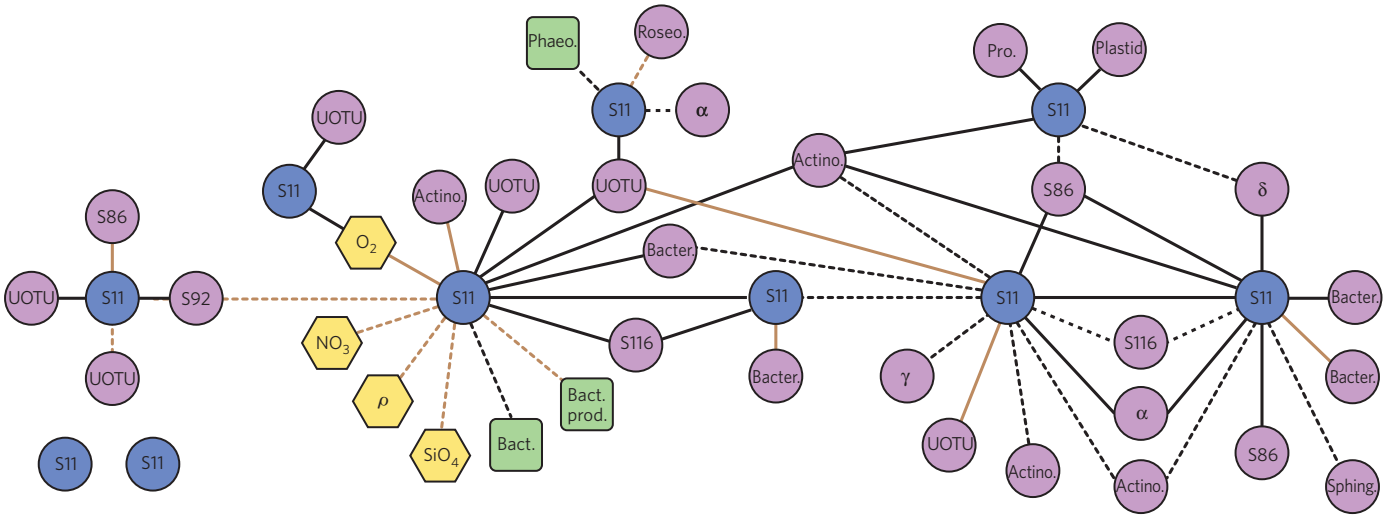


Figure 2 | Association networks. Systems-analysis tools can create association networks from field data on distributions, and these networks can have a time component if the measurements are a time series. This example from a 4.5-year time series off the California coast shows only the SAR11 bacteria and the organisms or parameters that directly correlate to them. There are ten discernible SAR11 subtypes (blue circles), but none of them shares the same combinations of correlations to abiotic (yellow hexagons) and biotic (green squares) environmental parameters or other individually distinguishable taxa (purple circles). Most of the SAR11 subtypes correlate to multiple bacterial uncharacterized operational taxonomic units (UOTUs) only, and few show correlations to environmental parameters directly. These relationships suggest that the ten SAR11 subtypes are candidates for different species, and that the differences between them are shown most clearly by their correlations to other bacteria and not by the relationships with environmental parameters. Solid lines indicate positive correlations. Dashed lines indicate negative correlations. Correlations with a 1-month time lag are indicated by brown lines. Data taken from ref. 46. α, α-proteobacteria; γ, γ-proteobacteria; δ, δ-proteobacteria; *ρ*, density; Actino., Actinobacteria; Bact., total bacteria; Bacter., Bacteroidetes; Bact. prod., bacterial production; NO_3, nitrate; O_2, oxygen; Phaeo., phaeopigments; Pro, *Prochlorococcus*; Roseo., *Roseobacter*; S, SAR; SiO_4, silicate; Sphing, *Sphingobacterium*.

advance of another, may help to sort out cause and effect.

To capture the important interactions that regulate system functions, network analysis should include all the important parameters in that system; the early networks, as shown in Fig. 2 with bacteria and some environmental parameters, show only the tip of the iceberg. Because microbial communities clearly include many interactions with protists, viruses and metazoans[51], all these organisms should ultimately be included in the analysis, otherwise important controlling factors will be missed. Considerable progress has been made in the molecular genetic characterization of marine protists[5,52–54] and viruses[55–57] (see also page 207). An initial analysis extending the interactions shown in Fig. 2 with Archaea and protists included shows that each bacterial OTU tends to have specific and multiple interactions with protist OTUs; sometimes the protist community seems more important than physicochemical parameters in determining the bacterial community structures (J. A. Steele, P. D. Countway, J. M Beman, L. Xia, J. Huang, P. D. Vigil, F. Sun, D. A. Caron, J. A. F., unpublished work). Network analysis can also be integrated with 'omics' studies (see page 200) for a more complete picture of organism functions in an environmental context. Holistic studies show potential interactions, and focused omics studies relate functions to organisms, illuminating the mechanism of the interactions.

Defining microbial species

Much debate has centred on the definition of microbial species, for theoretical and practical reasons[24,58,59]. What we call a species, and why, influences how we think about, study and understand organisms. It is relevant here to discuss how community structure can help evaluate species identities. Recent reports have demonstrated the value of characterizing the microenvironmental (ecological) preferences of organisms, such as season and size fraction for marine *Vibrio* spp.[60] or canyon microhabitat for *Bacillus* clades[59], to help verify the distinctiveness of closely related, genetically defined clusters that resemble species. I suggest that extending the analysis to include the co-occurrence of organisms can also help define ecological niches and lead to better ways of characterizing ecological species or microbial ecotypes. In other words, we should consider the biological environments of microorganisms, as well as their physicochemical environments, as important aspects of their niches. This is because many organisms, particularly microorganisms, live in microenvironments that are probably defined at least as much by the organisms around them as by physicochemical characteristics. Organisms create many aspects of the environment themselves (including subtle ones such as dilute nutrients), and their presence also provides a sort of monitoring and long-term historical integration of environmental conditions that are hard to measure directly, yet that may relate to the essence of a niche. An example would be to consider the ten SAR11 types in Fig. 2 as candidates for different ecological species or ecotypes. Obviously, more information than these correlations would be needed to confirm such distinctions (not easy in the case of organisms in their natural habitats), but networks provide a starting point and can hasten the development of descriptive or predictive models.

Predictable patterns of change

Multi-year monitoring by fingerprinting of SPOT samples reveals repeating patterns of bacterial community composition that are predictable not only on the basis of seasons (the month of occurrence), but also on the basis of environmental conditions such as temperature, salinity, chlorophyll and nutrient abundance, and viral and bacterial abundance[61]. Here, predictability refers to the statistical analysis of past data, but it suggests the possibility of forecasting changes in microbial communities. By expanding such approaches, we can start to consider how global change might alter the microbial landscape locally or globally.

The predictability of community composition from environmental parameters has some profound implications. It implies the presence of well-defined, probably narrow, niches for the predictable organisms. It also implies the related conclusion that the combinations of functional properties in each of the predictable organisms is unique (within these environments, at least), because redundancy or interchangeability between organisms would lead to random replacements that would interfere with the predictions. This non-redundancy is consistent with classical competitive exclusion or niche partitioning leading to one

organism in each niche. Predictability also implies a stable and consistent phenotype associated with the ARISA-defined genotype used to characterize members of the communities (the length of the spacer between 16S and 23S rRNA genes). This has implications for our interpretation of genetic data in general, as we discuss later.

Viral effects

Viruses are the most abundant biological entities in the sea, typically being an order of magnitude more abundant than bacteria. They are crucial components of marine systems, having major effects on ecological, biogeochemical and evolutionary processes[55,62–64] (see also page 207). They have left their genetic tracks widely in marine microbial genomes, for example in *Prochlorococcus*[65], and have short-term and long-term effects on the community structure of bacteria and protists[51]. Of particular relevance is the 'kill the winner' hypothesis[66], which posits that because virus infection is generally host specific and density dependent, abundant organisms are most susceptible to epidemic infection, and infection would reduce their abundance, leading to a succession of dominant organisms over time as one after another is attacked[62,66,67]. This is the prediction based on theory, but when we look at field data, the dominant OTUs often stay dominant for extended periods, certainly longer than this theory predicts[51]. One possible explanation is that OTUs are made up of multiple 'strains' with different virus sensitivities, and a closer examination would reveal a succession of different dominant strains. OTU definitions are typically based on slowly evolving genes that are generally not transferred (such as 16S rRNA), and virus resistance is often determined by more rapidly evolving or more easily transferred genes.

Linking community structure to genes and functions

Does the presence of an organism, an OTU or a functional gene tell us about a particular biogeochemical function? This is a fundamental question about the applicability of genetic studies to investigations of ecosystem function. Transcriptomics or proteomics will help (see page 200), but such studies are costly and the data are not as widely available as genetic information. The question arises because many genes are expressed only sometimes by the organism, such as the nitrogenase gene during nitrogen fixation[22]. In addition, different members of closely related groups of organisms — even well-defined species such as *Escherichia coli* — can have widely different genomes[24], so finding the species or OTU by a phylogenetic marker may not necessarily mean that other particular genes accompany it.

Core and flexible genomes

Intraspecific genomic consistency brings up the concept of 'core' and 'flexible' genomes of organism clusters, for example in the abundant marine cyanobacterium *Prochlorococcus*. Extensive study of numerous isolates of *Prochlorococcus*[65] has found that the genomes share more than 1,200 core genes (the 'core genome') that tend to perform functions of central housekeeping, such as DNA and protein synthesis. These core genes are thought to code for a functional cell, and phylogenetic trees constructed from them are generally congruent, implying little transfer between lineages[65]. These organisms also have many genes present only in some isolates. The 12 sequenced *Prochlorococcus* isolates collectively have about 4,000 flexible genes (there is usually a similar number of flexible and core genes in a given cell). Collectively called the 'flexible genome', these genes typically encode accessory functions such as the transport or use of certain non-essential nutrients, protection against damage from high light levels, and certain cell-surface modifications that probably relate to phage or predator resistance, but many have no known function. These flexible genes apparently control important aspects of the niche adaptations that define where the organisms can thrive. They are gained, lost or transferred much more than the core genes. The topic of core and flexible genomes is highly relevant for the interpretation of how community structure relates to function, because we use sequences or markers from the core genome to identify organisms, but genes in the flexible genome often define many of the functional characteristics.

Is the relationship between the core and accessory genome so variable that we cannot predict relevant environmental functions from core genome data such as 16S rRNA genes? It is too early to say whether the divergence between core and flexible genomes is a serious problem for interpreting community structure, but so far the data suggest that it is not a great problem. In *Prochlorococcus*, many genes coding for obvious niche-defining features such as optimal light levels and use of common nutrients seem to be shared among the most closely related strains[65], which is encouraging. Empirical data on distributions of six 16S-rRNA-based *Prochlorococcus* ecotypes, over wide areas of the ocean, suggested sensible distribution patterns that were consistent with stable phenotypes[68]. The 16S rRNA sequence differences between ecotypes were small, less than 3%, so OTUs were defined narrowly. The data on repeating patterns and predictability of bacterial community composition at the SPOT site[61] suggest that the relationship between phylogenetic markers and niche-defining phenotypes is usually fairly predictable. These field results imply that the marker genes (ARISA 16S–23S rRNA spacer length) were consistently associated with the part of the phenotype that made the organism predictable (its niche). They suggest that a particular core genome mostly has a consistent flexible genome. In the assemblage as a whole and over time, there seems to be a similar and consistent set of accessory genes associated with each core genome; that is, if there are multiple accessory genomes and phenotypes associated with a particular core genome, the proportions stay fairly constant. This further implies general stability of the community collective genome over ecological timescales. Studies currently under way will show whether these early encouraging results hold up.

Linking functions with genes

One goal of metagenomics (see pages 200 and 207) is to link particular functions to particular organisms. This requires the annotation of gene sequences with their functions, which can be a challenge, as shown by the example of proteorhodopsin, one of the most exciting early discoveries of metagenomics[69]. Initial analysis showed that this protein can act like a light-driven proton pump, allowing the cell to generate usable energy from light (phototrophy)[69]. But more recent results, including the lack of any clear growth benefit from light in most cultured organisms that have the gene, question whether phototrophy commonly occurs in the many organisms with this gene and whether the gene may have other functions as well[70]. So it may not be straightforward to interpret the function of a gene found in the environment, even when studied extensively. Linking genes with functions is probably best done by the integration of metagenomics, cultivation and field measurements.

Future prospects

New low-cost, high-throughput sequencing can greatly advance the analysis of marine microbial community structure, especially for measurements spread over space and time. Difficulties will remain, however, such as discovering the distributions of rare organisms. But this need not stop us following, modelling and eventually predicting the distributions of the majority of microorganisms and their activities, a critical aspect of understanding biogeochemical cycles in our changing world. ■

1. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
2. Pace, N. R., Stahl, D. A., Lane, D. L. & Olsen, G. J. The analysis of natural microbial populations by rRNA sequences. *Adv. Microb. Ecol.* **9,** 1–55 (1986).
3. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
4. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4,** e368 (2006).
5. Massana, R. & Pedrós-Alió, C. Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* **11,** 213–218 (2008).
6. Giovannoni, S. J. & Stingl, U. Molecular diversity and ecology of microbial plankton. *Nature* **437,** 343–348 (2005).
7. Baas Becking, L. G. M. *Geobiologie of Inleiding tot de Milieukunde* (Van Stockum & Zoon, 1934).
8. Pedrós-Alió, C. Marine microbial diversity: can it be determined? *Trends Microbiol.* **14,** 257–263 (2006).
9. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nature Rev. Microbiol.* **4,** 102–112 (2006).
This review combines ideas from microbial ecologists and 'macrobial' ecologists about how to include microorganisms in broader ecological theories and conceptual frameworks.

10. Foissner, W. Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool.* **45,** 111–136 (2006).
11. Ramette, A. & Tiedje, J. M. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* **53,** 197–207 (2007).
12. Fierer, N. in *Accessing Uncultivated Microorgansims: From the Environment to Organisms and Genomes and Back* (ed. Zengler, K.) 95–115 (ASM Press, 2008).
13. Fenchel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54,** 777–784 (2004).
14. Green, J. & Bohannan, B. J. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21,** 501–507 (2006).
15. Cottrell, M. T. & Kirchman, D. L. Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnol. Oceanogr.* **48,** 168–178 (2003).
16. Malmstrom, R. R., Cottrell, M. T., Elifantz, H. & Kirchman, D. L. Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl. Environ. Microbiol.* **71,** 2979–2986 (2005).
17. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl Acad. Sci. USA* **103,** 12115–12120 (2006).
18. Huber, J. A. *et al.* Microbial population structures in the deep marine biosphere. *Science* **318,** 97–100 (2007).
This is the second of two papers describing detailed analysis of microbial community structure in marine samples studied by a high-throughput sequencing technique.
19. Pommier, T. *et al.* Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* **16,** 867–880 (2007).
20. Brown, M. V., Schwalbach, M. S., Hewson, I. & Fuhrman, J. A. Coupling 16S-ITS rDNA clone libraries and ARISA to show marine microbial diversity: development and application to a time series. *Environ. Microbiol.* **7,** 1466–1479 (2005).
21. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276,** 1221–1229 (1997).
22. Zehr, J. P. & Ward, B. B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* **68,** 1015–1024 (2002).
23. Montoya, J. P. *et al.* High rates of N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430,** 1027–1031 (2004).
24. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56,** 457–487 (2002).
25. Staley, J. T. Biodiversity: are microbial species threatened? *Curr. Opin. Biotechnol.* **8,** 340–345 (1997).
26. Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B* **253,** 167–171 (1993).
27. Degnan, P. H., Lazarus, A. B., Brock, C. D. & Wernegreen, J. J. Host–symbiont stability and fast evolutionary rates in an ant–bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus blochmannia*. *Syst. Biol.* **53,** 95–110 (2004).
28. Brown, M. V. & Fuhrman, J. A. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat. Microb. Ecol.* **41,** 15–23 (2005).
29. Garcia-Martinez, J. & Rodriguez-Valera, F. Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol. Ecol.* **9,** 935–948 (2000).
30. Acinas, S. G. *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430,** 551–554 (2004).
31. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nature Rev. Microbiol.* **5,** 384–392 (2007).
This paper shows why the ongoing molecular revolution in microbial ecology needs to be driven by theory, and that the generality of established ecological theory must be tested using microbial systems.
32. Pianka, E. R. Latitudinal gradients in species diversity: a review of concepts. *Am. Nat.* **100,** 33–46 (1966).
33. Rosenweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
34. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163,** 192–211 (2004).
35. Hillebrand, H. Strength, slope and variability of marine latitudinal gradients. *Mar. Ecol. Prog. Ser.* **273,** 251–267 (2004).
36. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103,** 626–631 (2006).
37. Fuhrman, J. A. *et al.* A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl Acad. Sci. USA* **105,** 7774–7778 (2008).
38. Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. A taxa–area relationship for bacteria. *Nature* **432,** 750–753 (2004).
39. Woodcock, S. *et al.* Taxa–area relationships for microbes: the unsampled and the unseen. *Ecol. Lett.* **9,** 805–812 (2006).
40. Hewson, I., Steele, J. A., Capone, D. G. & Fuhrman, J. A. Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* **51,** 1274–1283 (2006).
41. Gotelli, N. J. & Graves, G. R. *Null Models in Ecology* (Smithsonian Institution Press, 1996).
42. Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* **83,** 2091–2096 (2002).
43. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **62,** 171–180 (2007).
44. Horner-Devine, M. C. *et al.* A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* **88,** 1345–1353 (2007).
45. Ruan, Q. S. *et al.* Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22,** 2532–2538 (2006).
46. Fuhrman, J. A. & Steele, J. A. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* **53,** 69–81 (2008).
47. Morris, R. M. *et al.* Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol. Oceanogr.* **50,** 1687–1696 (2005).
48. Montoya, J. M., Pimm, S. L. & Sole, R. V. Ecological networks and their fragility. *Nature* **442,** 259–264 (2006).
49. Long, R. A. & Azam, F. Antagonistic interactions among marine pelagic bacteria. *Appl. Environ. Microbiol.* **67,** 4975–4983 (2001).
50. Davidov, Y., Friedjung, A. & Jurkevitch, E. Structure analysis of a soil community of predatory bacteria using culture-dependent and culture-independent methods reveals a hitherto undetected diversity of *Bdellovibrio*-like organisms. *Environ. Microbiol.* **8,** 1667–1673 (2006).
51. Fuhrman, J. & Hagström, Å. in *Microbial Ecology of the Oceans* (ed. Kirchman, D.) 45–90 (Wiley, 2008).
52. Not, F., Gausling, R., Azam, F., Heidelberg, J. F. & Worden, A. Z. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* **9,** 1233–1252 (2007).
53. Massana, R. *et al.* Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.* **70,** 3528–3534 (2004).
54. Countway, P. D. *et al.* Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environ. Microbiol.* **9,** 1219–1232 (2007).
55. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5,** 801–812 (2007).
56. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4,** e368 (2006).
57. Bench, S. R. *et al.* Metagenomic characterization of Chesapeake bay virioplankton. *Appl. Environ. Microbiol.* **73,** 7629–7641 (2007).
58. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Rev. Microbiol.* **6,** 431–440 (2008).
This paper brings together many viewpoints on the definitions and nature of microbial species.
59. Koeppel, A. *et al.* Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl Acad. Sci. USA* **105,** 2504–2509 (2008).
60. Hunt, D. E. *et al.* Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320,** 1081–1085 (2008).
61. Fuhrman, J. A. *et al.* Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103,** 13104–13109 (2006).
This paper shows the predictability of marine microbial communities studied over a period of years, implying many features similar to familiar animal and plant systems.
62. Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64,** 69–114 (2000).
63. Fuhrman, J. A. Marine viruses: biogeochemical and ecological effects. *Nature* **399,** 541–548 (1999).
64. Suttle, C. A. Viruses in the sea. *Nature* **437,** 356–361 (2005).
65. Kettler, G. C. *et al.* Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* **3,** 2515–2528 (2007).
66. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13,** 19–27 (1997).
67. Fuhrman, J. A. & Suttle, C. A. Viruses in marine planktonic systems. *Oceanography* **6,** 51–63 (1993).
68. Johnson, Z. I. *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311,** 1737–1740 (2006).
69. Beja, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289,** 1902–1906 (2000).
70. Fuhrman, J. A., Schwalbach, M. S. & Stingl, U. Proteorhodopsins: an array of physiological roles? *Nature Rev. Microbiol.* **6,** 488–494 (2008).
71. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43 (2004).
72. Hewson, I. & Fuhrman, J. A. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Appl. Environ. Microbiol.* **70,** 3425–3433 (2004).
73. Quince, C., Curtis, T. P. & Sloan, W. T. The rational exploration of microbial diversity. *ISME J.* **2,** 997–1006 (2008).
74. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** e77 (2007).
75. Hewson, I., Steele, J. A., Capone, D. G. & Fuhrman, J. A. Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar. Ecol. Prog. Ser.* **311,** 67–77 (2006).
76. Franklin, M. P. *et al.* Bacterial diversity in the bacterioneuston (sea surface microlayer): the bacterioneuston through the looking glass. *Environ. Microbiol.* **7,** 723–736 (2005).
77. Fuhrman, J. A. in *Manual of Environmental Microbiology* 3rd edn (eds Hurst, C. J. *et al.*) Ch. 35 (ASM Press, 2007).
78. Ramette, A. Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62,** 142–160 (2007).
79. Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62,** 361–371 (2006).

Acknowledgements I thank Å. Hagström, F. Sun, J. Steele, I. Hewson, S. Naeem, J. Green and J. Brown for helping develop ideas presented here, and C. Chow, C. Horner-Devine and J. Cram for comments. Work in my laboratory was supported by the US National Science Foundation Microbial Observatories Program and grants 0527034, 0623575, 0648581, and 0703159, and by the University of Southern California Wrigley Institute for Environmental Studies.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (fuhrman@usc.edu).

NATURE|Vol 459|14 May 2009|doi:10.1038/nature08059

The microbial ocean from genomes to biomes

Edward F. DeLong[1]

Numerically, microbial species dominate the oceans, yet their population dynamics, metabolic complexity and synergistic interactions remain largely uncharted. A full understanding of life in the ocean requires more than knowledge of marine microbial taxa and their genome sequences. The latest experimental techniques and analytical approaches can provide a fresh perspective on the biological interactions within marine ecosystems, aiding in the construction of predictive models that can interrelate microbial dynamics with the biogeochemical matter and energy fluxes that make up the ocean ecosystem.

Just 40 years ago, the number of microorganisms in each millilitre of sea water was underestimated by a staggering three orders of magnitude. Astronauts may have been exploring the Moon, but most of the microbial life on Earth remained largely undiscovered. The situation changed dramatically in the late 1970s and early 1980s, when accurate estimates of total cell numbers in the sea became available. Over the next 25 years or so, local, regional and global estimates of microbial numbers, along with their bulk production and consumption rates in ocean surface waters, were quantified and mapped. These data provided increasingly accurate estimates of the total biomass of planktonic microorganisms and their turnover, enlarging their perceived role and significance in ocean food webs. Although this information was extremely useful, more specific data on the biology of planktonic Bacteria and Archaea have only recently become available, allowing us to address a new range of questions. Which taxa of marine Bacteria and Archaea are most dominant or biogeochemically important in particular ocean provinces or depth strata? What are the most common microbial metabolic pathways, and how do they vary within and between communities and environments? How do dynamic population shifts and species interactions shape the ecology and biogeochemistry of the seas?

Unlike eukaryotic plankton, which can often be taxonomically and metabolically categorized according to directly observable phenotypes, it has been more difficult to ascertain the core identities and physiological properties of planktonic Bacteria and Archaea. Recent advances in cultivation-independent metagenomics, in which DNA from the microbial community is collected, sequenced and analysed en masse, as well as new cultivation technologies, have had a dramatic influence on our knowledge of non-eukaryotic microorganisms. The integrated perspective provided by a combination of cultivation-independent phylogenetic surveys, microbial metagenomics and culture-based studies has delivered a more detailed understanding of microbial life in the sea. Here I discuss some of the contributions and synergy of metagenomics and the new cultivation approaches, focusing on recent advances achieved using these new techniques.

Phylogenetic surveys and model systems

One of the drivers for developing cultivation-independent approaches for the phylogenetic identification of microorganisms[1] was the recognition that only a small proportion of the microbial cells sampled from the environment can be readily cultivated using conventional techniques[2]. The development of ribosomal-RNA-based phylogenetic surveys in the 1980s led to less biased assessments of the distribution of uncultivated bacterial, archaeal and protistan phylotypes in natural populations[1]. The number of newly recognized bacterial and archaeal phylogenetic divisions has increased markedly. Indeed, in many habitats, some of the most abundant microbial phylotypes have no close relatives that have been cultured[3]. These and other results from cultivation-independent surveys have fundamentally changed our perspective on microbial phylogeny, evolution and ecology. These discoveries subsequently inspired more directed cultivation strategies, aimed at isolating some of the more environmentally abundant microbial phylotypes that had previously escaped cultivation[4–6].

Directed cultivation still has an important role in describing the nature and properties of marine Bacteria and Archaea. For example, the ocean's most abundant cyanobacterium, *Prochlorococcus*, which was first discovered by ship-board flow cytometry[7], was successfully cultivated soon after its discovery[8]. Isolates of *Prochlorococcus* now provide an environmentally relevant system for modelling the biology and ecology of planktonic cyanobacteria. Physiological characterization of *Prochlorococcus* genotypic variants led to the idea of 'ecotypes', which are highly related yet physiologically and genetically distinct populations that are adapted to different environmental conditions. An oceanographic survey of six *Prochlorococcus* ecotype variants in the Atlantic Ocean confirmed their distinct environmental distributions across broad environmental isoclines. *Prochlorococcus* isolates have also been used in detailed studies of phage diversity, host range, genome content, host–phage genetic exchange[9] and gene-expression dynamics[10]. The integration of *Prochlorococcus* lab-based physiological modelling and field-based surveys has also helped constrain and validate some computational ecosystem models that can successfully recapitulate known *Prochlorococcus* ecotype distributions in the environment[11], suggesting promising future directions in microbial oceanography.

The development of 'dilution to extinction' cultivation techniques[4] is another important advance aimed at culturing the new phylotypes discovered in rRNA-based environmental surveys. The basic approach involves preparing sterilized sea water, which is distributed into tissue-culture wells and subsequently inoculated with serially diluted bacterioplankton[6]. Growth in these low-density cultures is monitored by cell counting. These approaches have been hugely successful with respect to the recovery in pure culture of many dominant surface-water bacterioplankton[4–6,12].

[1]Departments of Biological Engineering and Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

As with any approach, however, there are practical limitations, such as uncertainties when using undefined and variable seawater media and the low probability of isolating rare organisms in the dilution-to-extinction approach. Indeed, the reasons why some predominant groups are readily cultivated, whereas others continue to resist cultivation, are still not well understood[4]. Nevertheless, the isolation and partial characterization of more representative bacterioplankton strains is having a major impact on our understanding of their genomic, phenotypic and physiological properties. The effects of these new approaches to cultivation are especially evident in the isolation of *Pelagibacter ubique*[12], a member of perhaps the most abundant bacterial group in the oceans. Isolates of *P. ubique* are now yielding fresh data on the phenotype[12,13], genome content[14], genetic variability[5,14] and physiology[15,16] of this major bacterioplankton taxon.

The cultivation of resident microorganisms is a valuable part of the drive to describe microbial processes in the environment, but it is not enough on its own. Although pure cultures provide readily manipulated models, there are fundamental limitations to their utility when it comes to inferring ecological processes. Some physicochemical variables can be well controlled in cultures, but patterns of temperature, pressure, pH, nutrient concentrations and redox balance, and their naturally occurring gradients, may sometimes be difficult to reproduce in the laboratory. Additionally, many microorganisms have evolved to interact closely with other organisms and are often engaged in obligatory symbiotic relationships. For these and other reasons, it is unreasonable to assume that pure-culture microbial models will be available for all the ecologically important microorganisms. Cultivation-independent phylogenetic and genomic surveys will continue to have an important role in describing uncultured microorganisms and their population genetics and biogeochemical and ecological interactions, which cannot be well studied or modelled in laboratory systems.

Microbial metagenomics and cultivation

For the purpose of this Review, 'metagenomics' is defined as the cultivation-independent genomic analysis of microbial assemblages or populations. Although still in its infancy, metagenomics has already contributed to our knowledge of genome structure, population diversity, gene content and the composition of naturally occurring microbial assemblages. In low-complexity populations, metagenomic studies have led to the assembly of almost complete genomes from the abundant genotypes[17] and have provided composite genomic representations of dominant populations[18,19]. Advances and improvements in sequencing technologies are propelling the field forward rapidly (Box 1). Despite the large data sets now available, high allelic variation in microbial populations, high species richness and a relatively even representation among species still render whole-genome assemblies of individual genotypes mostly impractical, given current sequencing and assembly technologies[20–22] (Box 2).

The coupling of metagenomics and culture-based approaches is particularly useful. Every methodology has its own shortcomings (see Box 2), but metagenomic surveys have already contributed significantly to our understanding of the microorganisms in the environment. For example, metagenomic data sets have allowed the directed enrichment and isolation of new isolates with specific and predicted functional and genetic properties[23]. In metagenomic surveys along environmental gradients, direct observations of gene distributions in the water column have revealed patterns of vertical stratification of functional genes, bacteriophage and other genetic properties, providing clues about the differential distribution of metabolic processes, phage–host interactions and evolutionary dynamics along the depth continuum[24]. A more recent survey using the latest pyrosequencing technologies compared more than 70 marine metagenomic data sets and revealed statistically significant differences in gene content among the nine major biomes compared[25]. In a recent dramatic example of cell-specific metagenomics, the genome content of an uncultivated nitrogen-fixing cyanobacterium population (UCYN-A) recovered by flow cytometry has been reported[26]. The genome sequences of the UCYN-A cell population revealed that these cyanobacteria, as expected, contained all the genes required for nitrogen fixation and all the components of photosystem I. The big surprise was that UCYN-A lacked the genes required for carbon dioxide fixation and oxygenic photosynthesis that are found in all other known free-living cyanobacteria[26]. The metagenomic data suggest that these cyanobacteria are not oxygen-generating photoautotrophs. This study provides an excellent example of how metagenomics can be used to identify the metabolic capabilities of uncultivated microbial phylotypes, a crucial goal in microbial ecology.

Metagenomic analyses of bacterial and archaeal populations have often presaged the later findings of culture-based studies. More specifically, metagenomic data have revealed unexpected phylogenetic and environmental distributions of genes and metabolisms. Early metagenomic studies, for example, revealed the unexpected presence of a bacteriorhodopsin-like photoprotein gene in an abundant marine bacterioplankton group (SAR86)[27]. Biophysical and functional characterization of the proteorhodopsin gene product confirmed its ability to function as a light-driven proton pump[27]. Later metagenomic surveys revealed the high abundance and global distribution of these rhodopsins in marine planktonic Bacteria and Archaea[20,21,28–35]. Subsequent genome sequencing of cultivated marine isolates then confirmed the widespread distribution of rhodopsin genes in many taxa of marine Bacteria[13,36,37]. Similarly, metagenomics revealed new types of aerobic, anoxygenic photosynthetic

Box 1 | Evolving genomic technologies

The range of genomic and metagenomic data now available for marine microorganisms is expanding rapidly for a variety of reasons. First, the acquisition of whole genome sequences from cultivated strains of microorganisms is becoming much faster and cheaper, so genome sequences are accumulating rapidly, with thousands now in the pipeline. With respect to marine microorganisms, hundreds of whole or draft bacterial and archaeal genome sequences are already available in public databases. In addition, nucleic-acid sequences recovered directly from total microbial assemblages are fast outstripping microbial whole-genome sequence data. The drivers for this include an increasing awareness of the usefulness of such data, a few major expeditions that have contributed large volumes of shotgun sequence data, and advancing technologies[60] that are making large amounts of sequence data readily available.

In addition to the size of metagenomic data sets, the heterogeneity of data types and environments sampled is also expanding dramatically. Original data sets mainly included Sanger-based shotgun sequence data of cloned DNA captured in small insert clone libraries (about 3 kilobase pairs, kbp) or longer genome fragments (40–100 kbp) in bacterial artificial chromosomes (BACs). More recently, pyrosequencing techniques[60] that do not require DNA clone libraries (eliminating the associated labour and cost overheads) have rapidly evolved from initial read lengths of 100 bp to 450 bp. Other next-generation technologies that involve sequencing by synthesis but generate very short reads (around 25 bp) may also prove useful in metagenomics, if sufficient long-read reference databases are available. On the horizon are technologies that will allow even higher-throughput, longer-read, single-molecule sequencing[61,62]. These advances will make a huge difference with respect to the amount of data that can be collected, as well as the bioinformatic infrastructure that will be required for analysis and synthesis to occur.

Single-cell genome sequencing using multiple displacement amplification (MDA) techniques coupled with new sequencing technologies also promises better genomic access to uncultivated or rare microorganisms[63–65], although significant challenges remain[64,65]. Chief among these are contamination problems associated with the 'extreme amplification' of large amounts of DNA from a single cell. Additionally, inherent mechanisms of the MDA reaction itself result in uneven amplification and coverage of even single, pure genotypes[65]. Partial draft genomes can be produced from single cells but currently not without extraordinary efforts to reduce contamination and to normalize for uneven coverage[63,66]. Nevertheless, incremental improvements in single-genome sequencing in the future are likely to allow the recovery of more partial draft genomes from as-yet-uncultivated Bacteria and Archaea. These are expected to both provide benefits to and derive benefit from the more traditional metagenomic approaches currently in common use.

Box 2 | Problems with metagenomic methods

The technical constraints of microbial sampling, changes in sequencing technologies and the sheer complexity and size of the data sets all present significant challenges for interpreting and comparing genomic data from microbial communities. Some of the larger challenges are discussed below.

There are numerous technical challenges associated with even the seemingly simple task of obtaining representative and reproducible samples. Sampling strategies are always context dependent and are influenced by the type of microbial community, its environment, the spatial scale sampled, the population density and the presence of contaminating substances. There are many relevant questions. Do the cells need to be purified away from a soil, sediment or rock matrix? To reduce sample complexity, will the cells be separated by size from larger eukaryotic species? Do the cells need to be concentrated before the DNA is extracted? These and other concerns about sampling are central to the interpretation of the resultant data sets.

The methods used to recover and sequence DNA from microbial communities are also critical. Past approaches using Sanger sequencing have predominantly relied on the cloning of individual DNA molecules. Cloning biases are well known, and in some cases specific genes[68] (as well as specific phylogenetic groups[69,70]) may be under-represented in genomic and metagenomic clone libraries. However, problems with such biases have been largely overcome by pyrosequencing[61] and other next-generation sequencing technologies that sidestep the need to clone individual DNA molecules.

Another problem relates to functional gene predictions and annotation. Even preliminary tasks of gene characterization, including calling open reading frames, identifying taxonomic origins and inferring functional properties, are non-trivial enterprises in analyses of metagenomic data sets. Complicating factors include short sequence read lengths, poor sequence quality, the absence of gene-linkage context, and having extremely large data sets and uneven coverage. Several strategies for metagenomic open-reading-frame prediction[22,71,72], phylogenetic assignment[73,74] and functional predictions[22,75,76] have recently been developed, and improvements and new approaches to these fundamental tasks continue to evolve. For example, a study combining homology searches and gene neighbourhood analyses succeeded in specific functional gene predictions for 76% of the 1.4 Mbp examined[77]. Such advances, alongside customized metagenomic databases[51–53], promise to improve current capabilities for gene identification and the annotation of metagenomic data sets.

Statistical approaches for the comparison of metagenomic data sets have only recently been applied, so their development is at an early stage. The size of the data sets, their heterogeneity and a lack of standardization for both metadata and gene descriptive data continue to present significant challenges for comparative analyses. Statistical approaches to examine gene distributions in the environment have so far included gene-enrichment probability estimates in three-way comparisons[75], bootstrap resampling methods that evaluate gene-abundance confidence intervals deviating from the median in pairwise sample comparisons[78], canonical discriminant analyses that identify the genes that most influence distributional variance[25], and canonical correlation analyses that interrelate metabolic-pathway occurrence with multiple environmental variables[79]. However, only highly disparate sample types have been the subject of much statistical scrutiny. It will be interesting to learn the sensitivity limits of such approaches, along more fine-scale taxonomic, spatial and temporal microbial community gradients, for example in the differences between the microbiomes of human individuals[44]. As the availability of data sets and comparable metadata fields continues to improve, quantitative statistical metagenomic comparisons are likely to increase in their utility and resolving power.

bacteria in marine plankton[38], an observation that was later confirmed by strain-isolation studies[39,40].

The predictive power of metagenomics was also demonstrated in the finding of genes associated with ammonia oxidation in Archaea, a character previously found in just a few bacterial groups. Two concurrent metagenomic studies[20,41] reported that a specific clade of Crenarchaeota seemed to have the genes diagnostic for chemolithotrophic ammonia oxidation. At about the same time, enrichment cultures using ammonia as the sole energy source and CO_2 as the sole carbon source yielded an ammonia-oxidizing crenarchaeal isolate[42]. Parallel metagenomic analyses of the genome sequence from an uncultured crenarchaeon extended previous studies beyond a single gene in the pathway and suggested specific functional differences between the archaeal and bacterial ammonia-oxidizing metabolic pathways[18,43]. In a very short time period, Archaea came to be recognized as potentially important contributors to a part of the nitrogen cycle previously thought to be regulated solely by Bacteria.

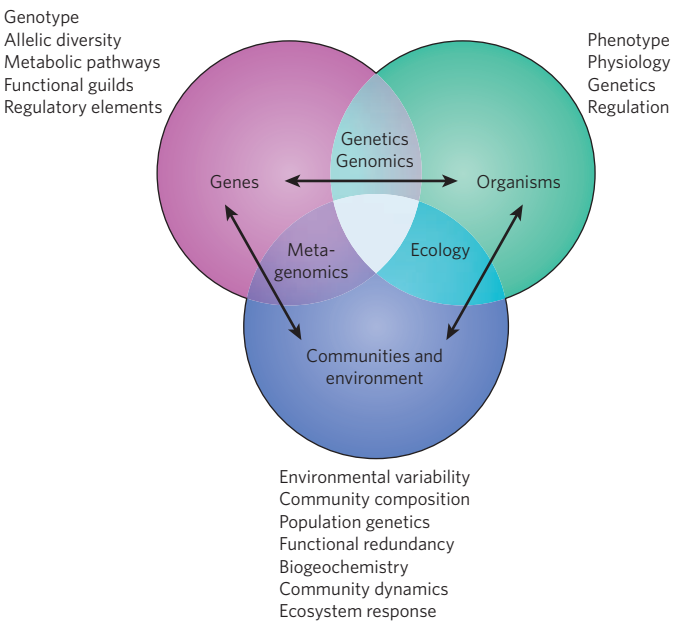


Figure 1 | The intersection of traditional disciplines and metagenomics. The pink, green and blue regions represent the fundamental elements of study: genes, organisms and the environment. Areas of investigation associated with each are indicated in the text. The intersections between the elements show the disciplinary overlaps: genetics/genomics, metagenomics and ecology. The pale blue area in the middle identifies the 'sweet spot' in which information from cultured-based studies, environmental studies and metagenomics can be integrated and modelled.

These and other examples have clearly indicated the value of integrating and comparing metagenomic and culture-based studies. Indeed, the deficiencies of each approach are largely compensated for by the strengths of the other. Phenotype, metabolism and physiology are mainly inferred from laboratory culture-based experiments, whereas detailed information on environmental distributions and ranges, population genetics, and community interactions and dynamics are best viewed through the lens of cultivation-independent strategies, including metagenomics. It is also clear that reference genome sequences from cultivated microorganisms greatly aid metagenomic studies. The integration of metagenomics, cultivation-based studies and environmental surveys leads to insights not previously open to microbiologists (Fig. 1), at the intersection of genes, organisms and the environment. More specifically, the integration of cultivation-dependent and cultivation-independent approaches partly bridges the gap between genomics, population genetics, biochemistry, physiology, biogeochemistry and ecology. Approaches that combine cultivation and metagenomic perspectives will undoubtedly be more common in future collaborative microbiological studies. Plans for human microbiome studies are a good case in point[44].

Nucleic-acid sequences as analytes in ecosystem studies

The development of metagenomic methods has helped to expand the repertoire of known microbial genes, their environmental distributions

and their allelic diversity. The associated bioinformatic analyses are useful for generating new hypotheses, but other methods are required to test and verify *in silico* hypotheses and conclusions in the real world. It is a long way from simply describing the naturally occurring microbial 'parts list' to understanding the functional properties, multi-scalar responses and interdependencies that connect microbial and abiotic ecosystem processes. New methods will be required to expand our understanding of how the microbial parts list ties in with microbial ecosystem dynamics. Experimental technologies that can leverage massively parallel sequencing technologies, or that can link information from pre-existing sequence data sets with experimental observations in natural assemblages, seem particularly promising.

Several approaches are available that have the potential to link DNA sequences found in the microbial community with specific microorganisms and their activities in the environment. One method uses the thymidine analogue 5-bromodeoxyuridine (BrdU) to tag actively growing substrate-responsive cells. The BrdU-labelled DNA is immuno-captured and subsequently sequenced to identify taxa and genes specific to a given experimental treatment[45]. Stable-isotope analyses also have significant potential for tracking specific microbial groups that incorporate labelled organic or inorganic compounds into living tissues. Stable-isotope tracers have been used to identify methanotrophic Archaea, to localize nitrogen-fixing symbionts in host tissues, and to verify autotrophic metabolism in planktonic Crenarchaeota. A novel approach that has the potential to link DNA sequence information directly to substrate-specific incorporation is stable-isotope probing, where nucleic acids labelled with a 'heavy' isotope are physically isolated by buoyant density centrifugation and subsequently sequenced[46].

The application of gene-expression technologies to track microbial sensing and responses in the environment is another exciting development. In this approach, bacterial and archaeal total RNA is extracted from microbial assemblages, converted to complementary DNA and sequenced (Fig. 2). Early studies began with the analysis of randomly primed cDNA clone libraries by Sanger-based capillary sequencing to survey abundant transcripts from a coastal seawater sample[47]. Advances such as pyrosequencing, which sidesteps the need for clone libraries, have allowed the analysis of larger data sets obtained from more rapidly collected, smaller-volume samples of marine bacterioplankton[48]. Pyrosequencing of both genomic DNA and cDNA from the same sample allows the normalization of transcript abundance to the corresponding gene copy number of the community's collective gene pool[48] (Figs 2, 3).

Early high-throughput, pyrosequence-based studies[48] of the transcriptome of planktonic microbial communities have led to several new insights. Not surprisingly, genes associated with the key metabolic pathways of open-ocean microbial species (including photosynthesis, carbon fixation and nitrogen acquisition) were found to be highly expressed in the photic zone at a depth of 75 m in the North Pacific Subtropical Gyre. Both genomic and transcriptomic data sets showed high coverage of some dominant community members, such as *Prochlorococcus,* with hypervariable genomic regions showing some of the highest transcript abundances. Many of the microbial community transcripts were similar to previously predicted genes found in ocean metagenomic surveys, but about half seemed to be unrelated to predicted protein sequences in available databases[48]. The transcriptomic data sets in such studies contain several categories of RNA, including rRNAs, messenger RNAs and small RNAs[49], some of which have an important role in regulating gene expression. Each of the molecular species recovered — rRNA, mRNA and small RNA — has the potential to shed light on the dynamics and variability of the phylogenetic composition, functional properties and regulation of natural microbial communities.

The application of transcriptomic methods to microbial communities is creating a new research agenda in which sequence data are the analytes in experimental field studies. This approach allows the measurement of gene expression in microbial assemblages, in microcosms, mesocosms or natural samples, as a function of environmental variability over time (Fig. 3). The environmental variation examined can be natural (for example, tracking changes in gene expression as a function of the daily cycle) or applied (for example, monitoring changes in gene expression following changes to nutrient levels). By tracking which genes are responsive to specific environmental perturbations, it should soon be possible to track environmental variations that are first observed as changes in gene expression but later may lead to shifts in community composition (Fig. 3). Quantifying the variability and kinetics of gene expression in natural assemblages has the

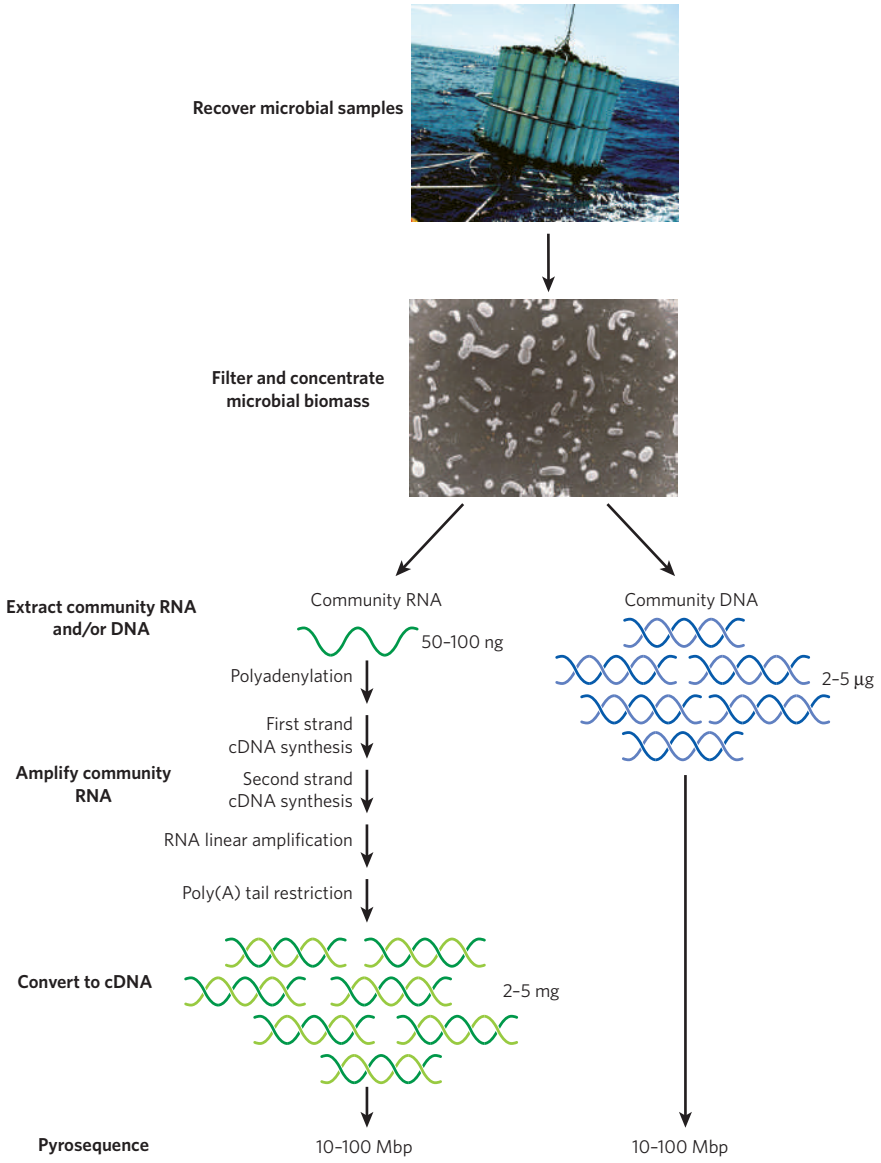


Figure 2 | Transcriptome sequencing protocol for marine microbial assemblages. Cells are collected and processed to produce genomic DNA, or cDNA from total RNA[48]; samples for RNA extraction are collected in smaller volumes (less than 1 litre) and filtered as rapidly as possible (about 10 min). After RNA amplification and conversion to cDNA, cDNA and genomic DNA from the same assemblage are sequenced and compared.

potential to provide a fresh perspective on microbial community dynamics. Can expression patterns provide clues to the functional properties of putative genes? What are the key community responses to environmental perturbation? What fundamental community-wide regulatory responses are common to different taxa? Are certain taxa or metabolic pathways more or less responsive to particular environmental changes? Are specific changes in gene expression indicative of changes in community composition? These and other questions can now be addressed more directly by applying these new experimental approaches.

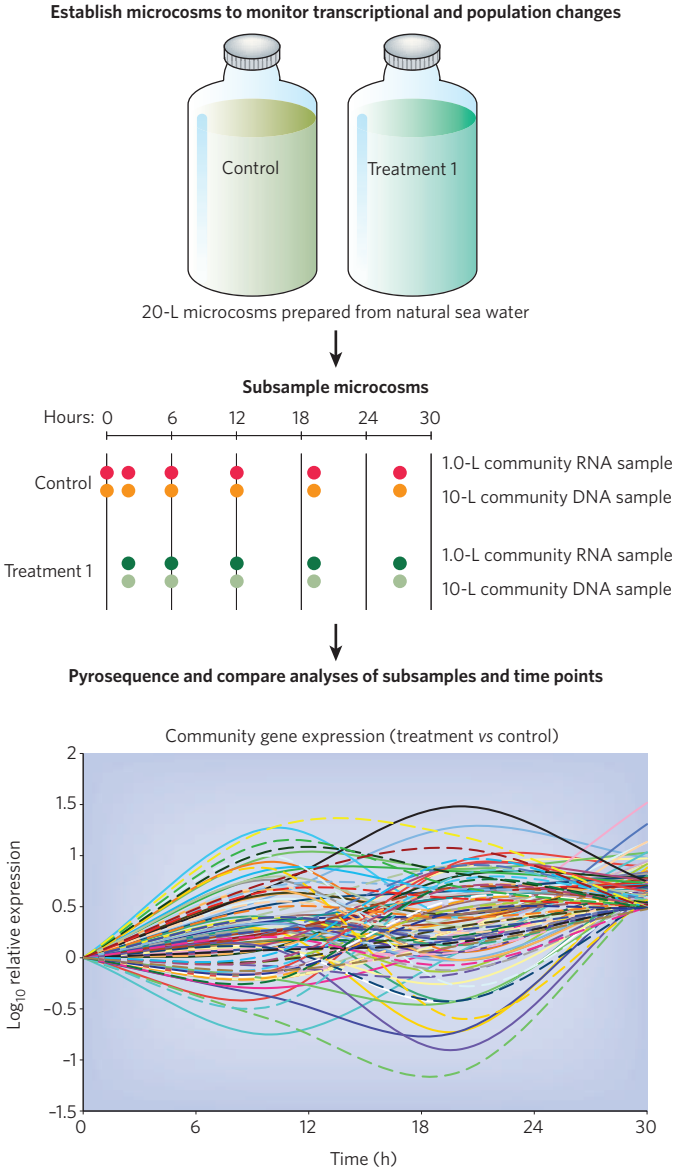


Figure 3 | Quantifying microbial responses to environmental variability using environmental transcriptomics. The experiments shown have been made possible by tandem metagenomic and 'metatranscriptomic' pyrosequencing (Fig. 2). Initially, microcosms containing aquatic microbial communities are established. The untreated sample is a control for intrinsic incubation effects, as well as natural daily variation in gene expression. Different experimental treatments could measure a variety of physical or environmental perturbations, including the effects of light, nutrients, temperature or anthropogenic compounds. Microbial-assemblage DNA and RNA subsamples are taken at various time points, subjected to pyrosequencing (see Fig. 2) and analysed and compared. Differential gene expression between control and treatment communities (bottom panel) is used to identify microbial responses to environmental perturbation. Coloured lines represent individual gene categories that are overexpressed or underexpressed relative to the control.

Information management from genes to ecosystems

One of the major challenges facing the emerging metagenomic and 'metatranscriptomic' studies is the sheer size of the data sets, and the methods and tools that are therefore needed to deal with them. Large data sets create challenges with respect to data management, computational resources, sampling and analytical strategies, and database architectures. It is encouraging that the research community has recognized the need to establish clear standards for the submission and reporting of data so that primary sequence data can be related across relevant environmental parameters. The Genomic Standards Consortium (http://gensc.org) is promoting schemes reminiscent of the MIAME standards for microarray data (http://www.mged.org/Workgroups/MIAME/miame.html). These would capture metadata associated with genomes (minimum information about a genome sequence) and metagenomic data (minimum information about a metagenome sequence)[50]. For comparative analyses of archived data sets, such metadata field standardization and reporting will be critical.

We are entering a new era in microbial ecology and biology in which experimental high-throughput sequencing data will increasingly be analysed (Fig. 3). The coordination of experimental reports from such studies will be important, and MIAME-like standards for such reporting (minimum information about a high-throughput sequencing experiment) have recently been proposed (http://www.mged.org/minseqe). Even simple annotation, archiving and accessing of sequence-data types and experiments, along with associated and relevant metadata, pose serious challenges for the biological community. These challenges are being addressed by the development of new metagenomic databases[51–53], analytical strategies and statistical approaches (Box 2).

Efficient bioinformatics management and analytical practices will not be a panacea for the larger challenge of describing microbial biology at an ecosystem level. There is still a mismatch with respect to the integration of 'bottom up', reductionist molecular, approaches with 'top down', integrative ecosystems, analyses. Molecular data sets are often gathered in massively parallel ways, but acquiring equivalently dense physiological and biogeochemical process data[54] is not currently as feasible. This 'impedance mismatch' (the inability of one system to accommodate input from another system's output) is one of the larger hurdles that must be overcome in the quest for more realistic integrative analyses that interrelate data sets spanning from genomes to biomes.

The road ahead

The microbial parts list of the genes and genomes in metagenomic data sets is growing rapidly, but work to understand their functional and ecological relevance is proceeding more slowly. DNA sequence data and bioinformatic analyses fall short of describing which gene suites are being expressed, and which metabolic pathways are being used, in any given environmental context. A large number of hypothetical proteins that have been identified may be ecologically important but have functions that remain unknown. How do community composition, gene content and variability influence biogeochemical function, turnover rates and ecosystem processes? How important are functional redundancy and allelic diversity to community function and stability? How does the process of succession play out, from the initial environmental change to shifts in microbial community composition? Can we predict the probability of lateral gene transfer and gene fixation for particular functional properties or gene categories? Can suites of genes and their variability be correlated with larger-scale biogeochemical and ecological patterns and processes? Can we determine the functional properties and roles of as-yet-uncharacterized proteins that share little or no homology with functionally annotated proteins? How representative are the activities and responses of microbial isolates in the laboratory, with respect to their physiological and metabolic behaviour in the environment? Fresh approaches will be required to address these and other questions that are currently being raised.

We need to develop and explore new strategies to bridge the gaps between microbial genomics, metagenomics, biochemistry, physiology, population genetics, biogeochemistry, oceanography and ecosystem

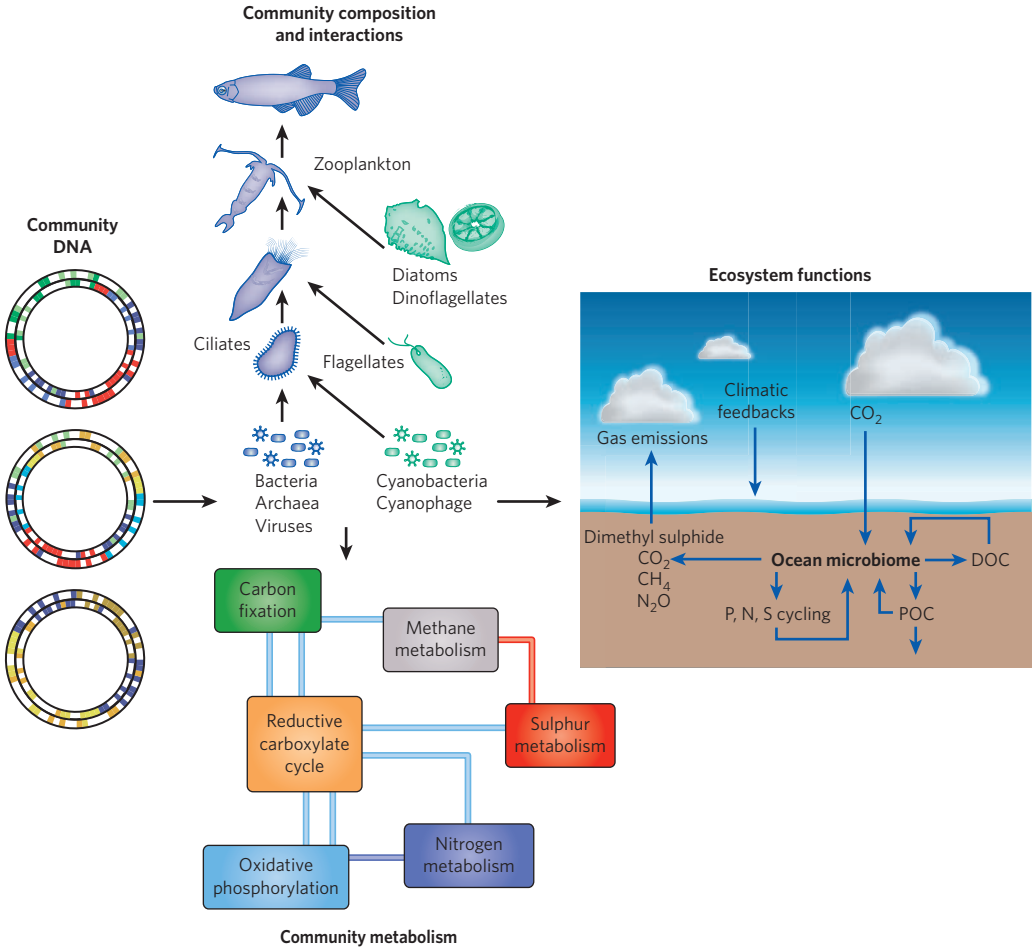


Figure 4 | The network instructions encoded in microbial genomes drive ecosystem processes. This schema shows hypothetical linkages between the genomic information of the microbial assemblage and the collective ecological interactions and community metabolism that in part regulate and sustain biogeochemical and ecosystem processes. Each DNA circle in the left panel represents a genome derived from a marine bacterioplankton species. Co-occurring microorganisms that inhabit the same environment collectively form the pool of genes sampled in metagenomic studies. These instructions modulate community interactions, metabolism and ecosystem function. DOC, dissolved organic carbon; POC, particulate organic carbon.

biology. Integrative and interdisciplinary interactions will be key to future studies because microbial diversity, metabolism and biogeochemistry are all intertwined over multiple temporal and spatial scales. One central hypothesis that drives metagenomics is that the network instructions for metabolic processes, biogeochemical function and ecological interactions are encoded in the collective microbial genomes and expressed in response to environmental variability. These network instructions are eventually expressed as the biological drivers of ecosystem processes (Fig. 4).

Microbial metabolic diversity and environmental variation together lead to changes in biological matter and energy flux. Time series[55] and mesocosm studies[56] are being used to investigate how microorganisms and their activities co-vary with environmental change. Efforts to integrate microbial diversity and process data with quantitative models that incorporate physical oceanography and biogeochemistry are still in their infancy[11,24,54,56–59]. Momentum is building, however, and direct observations of microbial diversity, variability and processes will soon inform models that will in turn inform and direct further field-oriented surveys, experiments and measurements. Observation, experiment and theory can together provide, verify and integrate information from genomics, metagenomics, microbial physiology, biogeochemistry and ecology, creating a clearer picture of emergent properties in the microbial systems that drive energy and matter flux in ocean ecosystems. The challenges to integrating work across disciplinary and conceptual boundaries are formidable, but the need for a more interdisciplinary understanding of the microbial ocean is clear. The reward will be a greatly improved qualitative and quantitative perspective on the living ocean system, from genomes to biomes. ■

1. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
2. Staley, J. T. & Konopka, A. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* **39,** 321–346 (1985).
3. Rappe, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57,** 369–394 (2003).
4. Giovannoni, S. & Stingl, U. The importance of culturing bacterioplankton in the 'omics' age. *Nature Rev. Microbiol.* **5,** 820–826 (2007).
5. Stingl, U., Tripp, H. J. & Giovannoni, S. J. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J.* **1,** 361–371 (2007).
6. Connon, S. A. & Giovannoni, S. J. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **68,** 3878–3885 (2002).
   This is the first report of a dilution-to-extinction cultivation approach that was successful in isolating a wide variety of the predominant marine bacterioplankton types.
7. Chisholm, S. W. *et al.* A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone. *Nature* **334,** 340–343 (1988).
8. Chisholm, S. W. *et al. Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch. Microbiol.* **157,** 297–300 (1992).
9. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424,** 1047–1051 (2003).
10. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449,** 83–86 (2007).
11. Follows, M. J., Dutkiewicz, S., Grant, S. & Chisholm, S. W. Emergent biogeography of microbial communities in a model ocean. *Science* **315,** 1843–1846 (2007).
12. Rappe, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418,** 630–633 (2002).
    This paper reports the first isolation in pure culture of a strain from the SAR11 clade, a representative of one of the most abundant bacterial groups in marine plankton.
13. Giovannoni, S. J. *et al.* Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438,** 82–85 (2005).

14. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309,** 1242–1245 (2005).
15. Tripp, H. J. *et al.* SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452,** 741–744 (2008).
16. Tripp, H. J. *et al.* Unique glycine-activated riboswitch linked to glycine–serine auxotrophy in SAR11. *Environ. Microbiol.* **11,** 230–238 (2009).
17. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43 (2004).
18. Hallam, S. J. *et al.* Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl Acad. Sci. USA* **103,** 18296–18301 (2006).
19. Allen, E. E. *et al.* Genome dynamics in a natural archaeal population. *Proc. Natl Acad. Sci. USA* **104,** 1883-1888 (2007).
20. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
21. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** e77 (2007).
22. Yooseph, S. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5,** e16 (2007).
This was the first attempt to identify protein-family clusters globally, based on public data combined with some 6 million newly predicted peptides from a metagenomic sampling of surface-water marine bacterioplankton.
23. Tyson, G. W. *et al.* Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* **71,** 6319–6324 (2005).
24. DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311,** 496–503 (2006).
25. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452,** 629–632 (2008).
26. Zehr, J. P. *et al.* Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. *Science* **322,** 1110–1112 (2008).
27. Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289,** 1902–1906 (2000).
This paper reports the first observation and characterization of ion-pumping rhodopsins in the domain Bacteria, a discovery enabled by metagenomic sampling and analyses.
28. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411,** 786–789 (2001).
29. Sabehi, G. *et al.* Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* **5,** 842–849 (2003).
30. de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100,** 12830–12835 (2003).
31. Sabehi, G., Beja, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6,** 903–910 (2004).
32. Sabehi, G. *et al.* New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.* **3,** e273 (2005).
33. Frigaard, N. U., Martinez, A., Mincer, T. J. & DeLong, E. F. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439,** 847–850 (2006).
34. McCarren, J. & DeLong, E. F. Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.* **9,** 846–858 (2007).
35. Martinez, A., Bradley, A. S., Waldbauer, J. R., Summons, R. E. & DeLong, E. F. Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl Acad. Sci. USA* **104,** 5590–5595 (2007).
36. Gomez-Consarnau, L. *et al.* Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445,** 210–213 (2007).
37. Stingl, U., Desiderio, R. A., Cho, J. C., Vergin, K. L. & Giovannoni, S. J. The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl. Environ. Microbiol.* **73,** 2290–2296 (2007).
38. Béjà, O. *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415,** 630–633 (2002).
39. Cho, J. C. *et al.* Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria. *Environ. Microbiol.* **9,** 1456–1463 (2007).
40. Fuchs, B. M. *et al.* Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **104,** 2891–2896 (2007).
41. Treusch, A. H. *et al.* Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. *Environ. Microbiol.* **7,** 1985–1995 (2005).
This is the first report presenting definitive metagenomic evidence for the occurrence of nitrification-associated genes (encoding ammonia monooxygenase subunits A and B) in non-thermophilic Crenarchaeota.
42. Konneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437,** 543–546 (2005).
This is the first report to prove the existence of chemolithoautotrophic, ammonia-oxidizing, non-thermophilic Crenarchaeota by their isolation in pure culture.
43. Hallam, S. J. *et al.* Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol.* **4,** e95 (2006).
44. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449,** 804–810 (2007).
45. Mou, X., Hodson, R. E. & Moran, M. A. Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. *Environ. Microbiol.* **9,** 2025–2037 (2007).
46. Neufeld, J. D., Wagner, M. & Murrell, J. C. Who eats what, where and when? Isotope-labelling experiments are coming of age. *ISME J.* **1,** 103–110 (2007).
47. Poretsky, R. S. *et al.* Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* **71,** 4121–4126 (2005).
48. Frias-Lopez, J. *et al.* Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA* **105,** 3805–3810 (2008).
This is the first report of a directed, massively parallel approach that sequenced a sample of the transcriptome in planktonic microbial assemblages by using pyrosequencing.
49. Shi, Y., Tyson, G. W. & DeLong, E. F. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459,** 266–269 (2009).
50. Field, D. *et al.* The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnol.* **26,** 541–547 (2008).
51. Markowitz, V. M. *et al.* IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36,** D534–D538 (2008).
52. Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P. & Frazier, M. CAMERA: a community resource for metagenomics. *PLoS Biol.* **5,** e75 (2007).
53. Meyer, F. *et al.* The metagenomics RAST server — a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9,** 386 (2008).
54. Anderson, R. *et al.* A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS). *Ambio* **10,** 4–30 (2001).
55. Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: the domain shift hypothesis. *Deep-Sea Res. II* **48,** 1449–1470 (2001).
56. Karl, D. M. Microbial oceanography: paradigms, processes and promise. *Nature Rev. Microbiol.* **5,** 759–769 (2007).
57. DeLong, E. F. Towards microbial systems science: integrating microbial perspective, from genomes to biomes. *Environ. Microbiol.* **4,** 9–10 (2002).
58. Doney, S., Abbott, M., Cullen, J., Kar, I. D. & Rothstein, L. From genes to ecosystems: the ocean's new frontier. *Front. Ecol. Environ.* **2,** 457–466 (2004).
59. Fuhrman, J. A. *et al.* Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103,** 13104–13109 (2006).
60. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
61. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).
62. Korlach, J. *et al.* Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc. Natl Acad. Sci. USA* **105,** 1176–1181 (2008).
63. Binga, E. K., Lasken, R. S. & Neufeld, J. D. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J.* **2,** 233–241 (2008).
64. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. & Lasken, R. S. Genomic sequencing of single microbial cells from environmental samples. *Curr. Opin. Microbiol.* **11,** 198–204 (2008).
65. Lasken, R. S. & Stockwell, T. B. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol.* **7,** 19 (2007).
66. Marcy, Y. *et al.* Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl Acad. Sci. USA* **104,** 11889–11894 (2007).
67. Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318,** 1449–1452 (2007).
68. Béjà , O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2,** 516–529 (2000).
69. Pham, V. D., Konstantinidis, K. T., Palden, T. & DeLong, E. F. Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ. Microbiol.* **10,** 2313–2330 (2008).
70. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34,** 5623–5630 (2006).
71. Krause, L. *et al.* Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics* **22,** e281–e289 (2006).
72. Krause, L. *et al.* Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* **36,** 2230–2239 (2008).
73. von Mering, C. *et al.* Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315,** 1126–1130 (2007).
74. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308,** 554–557 (2005).
75. Dalevi, D. *et al.* Annotation of metagenome short reads using proxygenes. *Bioinformatics* **24,** i7–i13 (2008).
76. Harrington, E. D. *et al.* Quantitative assessment of protein function prediction from metagenomics shotgun sequences. *Proc. Natl Acad. Sci. USA* **104,** 13913–13918 (2007).
77. Rodriguez-Brito, B., Rohwer, F. & Edwards, R. A. An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7,** 162 (2006).
78. Gianoulis, T. A. *et al.* Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc. Natl Acad. Sci. USA* **106,** 1374–1379 (2009).
Canonical correlation analyses identified metabolic pathways in metagenomic data sets that maximally co-varied with multiple environmental variables, revealing co-variation of amino-acid transport and cofactor synthesis across communities and environments.

Acknowledgements I thank my current and former students, colleagues and co-workers for sharing their ideas, insights, enthusiasm and inspiration. Work in my laboratory is supported by grants from the US National Science Foundation, the US Department of Energy, the Gordon and Betty Moore Foundation and the Agouron Institute. This article is a contribution from the NSF Science and Technology Center, and the Center for Microbial Oceanography: Research and Education.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (delong@mit.edu).

Viruses manipulate the marine environment

Forest Rohwer[1] & Rebecca Vega Thurber[1,2]

Marine viruses affect Bacteria, Archaea and eukaryotic organisms and are major components of the marine food web. Most studies have focused on their role as predators and parasites, but many of the interactions between marine viruses and their hosts are much more complicated. A series of recent studies has shown that viruses have the ability to manipulate the life histories and evolution of their hosts in remarkable ways, challenging our understanding of this almost invisible world.

Marine virology has traditionally focused on two areas: viruses as pathogens of aquatic organisms, and phage-driven dynamics of the marine microbial food web. Both of these influence global biogeochemistry and host evolution, and the former also has important economic and conservation implications. For example, two common marine viral diseases, sea-turtle fibropapillomatosis and shrimp white spot syndrome, endanger protected marine species and the financial stability of the aquaculture industry. Although marine virology is about 70 years old (Box 1), it has experienced a recent surge in interest[1,2], largely thanks to methodological advances.

One of the main areas of study in the past few years has been the extent of viral diversity in the marine environment. Diversity has been hard to measure because viruses do not have a universally conserved gene like the ribosomal DNA genes in cellular organisms, and because most viral hosts are difficult to culture. To circumvent these difficulties, whole viral communities have been isolated and analysed using pulsed-field gel electrophoresis or shotgun sequencing[3,4]. Shotgun sequencing led to the rise of marine viral metagenomics, which has shown that viruses are exceptionally diverse: there are more than 5,000 viral genotypes or species in 100 litres of sea water, and up to 1 million species in 1 kg of marine sediment[5,6]. Marine viral metagenomes, or 'viromes', collected from across the world have shown that viral species are globally distributed (everything is everywhere) but that the relative abundance of each species is restricted by local selection[7,8]. These studies have also shown that viral functional diversity, and its potential use for host adaptation, has been vastly underestimated[9].

Marine virology is now poised to move away from bulk measurements of predation and biodiversity towards the detailed analysis of evolution and ecology. In this Review, we show how marine viruses can affect their hosts and environments in startling ways. From the global transfer of niche adaptation genes to modifications of the ontogeny and ecology of marine organisms, it has become clear that the marine virome is a master of manipulation.

Virally encoded host genes

Phage, and to a lesser degree eukaryotic viruses[10], are known to carry and transfer a variety of host genes[11]. Most studies of this phenomenon have focused on the negative effects of viruses modifying their host's physiology. However, viral infections can augment the metabolism, immunity, distribution and evolution of their hosts in many unexpected and potentially positive ways (Fig. 1).

Consider the cyanobacterial genera *Synechococcus* and *Prochlorococcus*, which together account for about 25% of global photosynthesis[12]. Sequencing of the marine viral cyanophages that infect these primary producers showed that genes involved in photosynthesis are commonly carried in phage genomes[13]. These genes include the high-light-inducible (*hli*) gene, as well as *psbA* and *psbD*, which encode the photosystem II (PSII) core reaction-centre proteins D1 and D2, respectively[14] (Table 1). The D1 protein is of particular interest because it is the most labile protein in PSII and the most likely to be rate limiting. During the lytic cycle, most of the host's transcription and translation is shut down by phage. Because phage must maintain the proton motive force if they are to lyse the host, they need to prolong photosynthesis during the infection cycle. The cyanophage-encoded D1 proteins are expressed during the infection cycle, countering the virally induced decline in host gene expression[15]. It is thought that by encoding *psbA* and other genes involved in photosynthesis, phage generate the energy necessary for viral production.

One consequence of cyanophage carrying *psbA* genes is the horizontal gene transfer of photosynthetic genetic elements between hosts (Fig. 1). *Prochlorococcus* has specific ecotypes that live in different parts of the water column[16] and are tuned to the different light and nutrient regimes found there. Given the prevalence of phage-encoded photosynthesis proteins and the occurrence of recombination between phage and host genes, phage populations are expected to serve as gene reservoirs that change the ecological niches of the host[17]. Several lines of evidence support this hypothesis. First, phage *psbA* genes are undergoing independent selection from host *psbA*, and there has clearly been exchange of phage *psbA* between hosts[18]. Second, metagenomic analyses have routinely identified large numbers of *psbA* genes in viral fractions and associated with viral-like open reading frames. It has been estimated that about 60% of the *psbA* genes in the marine environment for which an origin could be identified were actually from phage[19]. A rough calculation suggests that some 10% of total global photosynthesis could be carried out as a result of *psbA* genes originally from phage.

Transformation events also mediate one of the most dramatic effects of phage on their hosts: the switch from symbiont or benign microorganism to pathogen. The best-known marine example occurs in *Vibrio cholerae*, a common near-shore bacterium that is normally harmless but becomes one of humanity's greatest scourges by incorporating phage cholera toxin (CTX) genes[20]. Large-scale metagenomics has shown that viruses contain high numbers of virulence genes (Table 1), including some that facilitate antibiotic resistance, toxicity, host adhesion and host invasion. Bacteria that take up these genes extend their ecological niches, although this ultimately has a negative impact on humans.

In addition to virulence genes, marine viromes contain many genes that are involved in unanticipated metabolic and functional pathways. Comparisons of paired microbial and viral fractions (microbiomes and

[1]Department of Biology, San Diego State University, San Diego, California 92182, USA. [2]Department of Biological Sciences, Florida International University, 3000 NE 151st street, North Miami, Florida 33181, USA.

viromes, respectively) show that the relative frequency of respiration genes is lower in the viromes, whereas genes involved in nucleic-acid metabolism are more abundant[9]. Less expected was the observation that microbiomes and viromes carry almost equal frequencies of metabolic genes involved in carbohydrate and protein metabolism. Totally unexpected was the finding that genes involved in vitamin and cofactor synthesis, stress-response genes such as those encoding chaperones, and genes associated with bacterial motility and chemotaxis were more common in viromes than in their corresponding microbiomes[9].

Viromes as novel gene banks

Viromes are good hunting grounds for unique host-adaptation genes, as shown in a recent metagenomic study of phage from deep-sea hydrothermal vents. The abundance of viral particles was found to be higher in the diffuse flow, a region where cold sea water mixes with warm fluids from hydrothermal vents, than in the surrounding sea water[21]. Both this observation and the taxonomic make-up of the viromes suggest that temperate prophages are being induced in the diffuse flow. Only about 25% of sequences from the vent viromes had any significant similarity to sequences in the GenBank database. This high abundance of novel sequence suggests that these deep-sea viral communities could be a store of genes that may be involved in microbial adaptation to the high pressures, high temperatures and high concentrations of inorganic chemicals (such as sulphides, iron, salt and calcium) found in vent systems.

Generalized transducing agents

If viromes serve as reservoirs of genes, then determining the rate of exchange between viruses and their hosts is important. One study found a high rate of transduction in the marine environment[22]. Extrapolation of these data suggests that as many as 10^{24} genes are moved by transduction from virus to host each year in the world's oceans. However, the actual amount of transduction by marine viruses and viral-like entities may be much greater than previously thought because of the action of generalized transducing agents (GTAs)[23]. GTA particles are similar in morphology to phage, but they are smaller (with a head diameter of 30–50 nm) and contain a smaller amount (about 4 kilobases, kb) of DNA. What makes GTAs unique is that they only carry host DNA, which is injected into a recipient[24], providing an efficient form of transduction.

GTAs were originally identified in the bacterium *Rhodobacter capsulatus* but have now been found in a variety of bacteria (including Spirochaetaceae and Proteobacteria) and archaeal organisms (including *Methanococcus*). The GTAs found in α-proteobacteria, such as the Rhodobacterales, have been shown to be vertically transmitted and to have evolved from a single common ancestor, and they probably arose before the diversification of bacterial phyla[24]. Because Rhodobacterales are extremely abundant in the ocean, the transmission of such genetic agents is likely to have significant consequences for marine microbial ecology. A recent study has shown that genes encoding GTAs are found in most marine systems[25]. The same study also showed that these GTAs are produced by marine bacteria and move genes between species of α-proteobacteria. These observations suggest that GTA-related gene swapping may contribute to the niche partitioning of closely related species in the ocean.

Gene swapping between domains and ecosystems

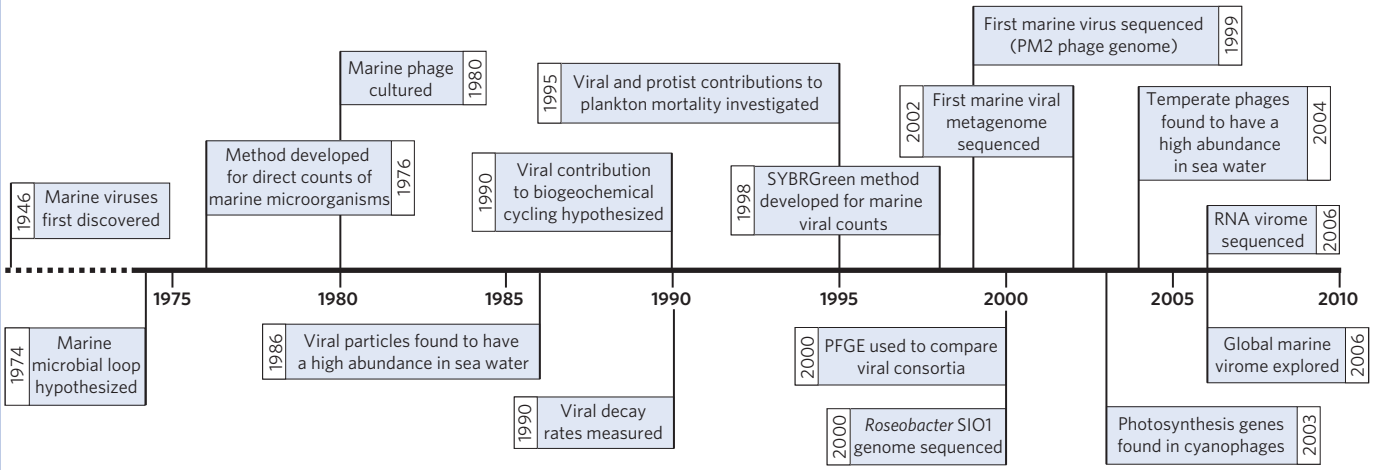
Known phage and GTAs have relatively restricted host ranges, limiting the rates by which genes can move from one host to another by these mechanisms. No known virus routinely moves between the three domains of life. However, viral-like particles in sea water and hot springs have been shown to transfer genes between Archaea, Bacteria and Eukarya[26,27]. The exact nature of these particles is still being analysed, but this finding opens up a new realm of horizontal gene transfer.

Viruses not only move genetic material from one organism to another, but from one ecosystem to another. Several phage sequences have been found to be spread ubiquitously through the biosphere[28,29]. There is also evidence that phage from one environment can successfully infect and replicate in marine microorganisms from unrelated environments[30]. These results support the hypothesis that viruses can move throughout the world and contribute to a global genetic pool. It may be the case that although local viral diversity is very high, total global diversity is limited by the worldwide movement of virions.

Viral manipulation of viruses

Acanthamoeba spp. are common protists found in soil and fresh and salt water[31]. They primarily graze on microorganisms, and some species

Box 1 | Highlights in marine virology research



Bacterial viruses (phage) in sea water were first observed in the first half of the last century[59,60], although their presence remained unexplained until Lawrence Pomeroy hypothesized the 'marine microbial loop'[61] in 1974. In 1979, Francisco Torrella and Richard Morita discovered that marine viral particles were particularly abundant (10^4 per millilitre) and morphologically similar to phage[62], and phage from marine bacteria were soon cultured[63]. In the 1990s, much was learned about the genetic diversity of marine phage and eukaryotic viruses and their importance to the ecology of the marine plankton community. Numerous studies demonstrated the contribution of viruses and protists to global biogeochemical cycling arising from the lysis of plankton[4,39,64–66]. The first marine viral genomes were then sequenced[57,67], and genomics and metagenomics have since been used to characterize the diversity of both RNA viruses[68,69] and DNA viruses[7,70] in sea water, along with their effects on host physiology and ecology[71–73]. The timeline shown here (not to scale) lists the main events in marine virology research. PFGE, pulsed-field gel electrophoresis.

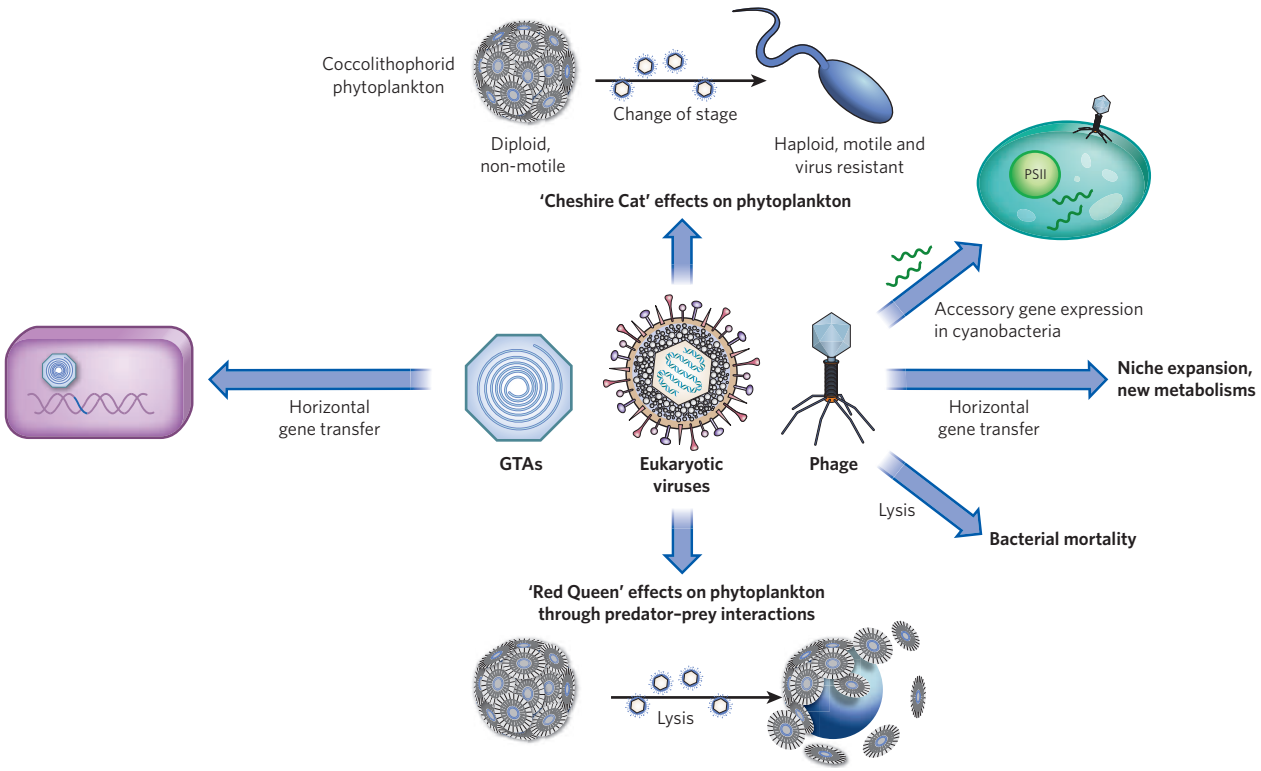


Figure 1 | Effects of marine viruses on their hosts. Marine viruses and viral-like entities, including eukaryotic viruses, phage and generalized transducing agents (GTAs), can have various effects on host cells. When a phage infects its bacterial host cell, it can either kill the cell (lysis), or transfer genetic material obtained from a previous host (horizontal gene transfer) or from its own genome (accessory genes expression). The transferred genes can allow a cell to expand into different niches (for example, through the activation of photosynthetic genes, changing the life cycle of biogeochemically important phytoplankton such as cyanobacteria and coccolithophorids). Similarly, small viral-like particles known as GTAs can transfer genes between marine organisms. Two possible scenarios have been proposed for viral effects on cells. In the 'Red Queen' effect, the virus and cell are locked in an evolutionary 'arms race', such that they continue to evolve mechanisms of resistance to each other until eventually the virus causes the host cell (illustrated here by a coccolithophorid) to die. In the 'Cheshire Cat' hypothesis, however, the coccolithophorid simply moves from its diploid, non-mobile stage to a motile, haploid stage, thereby evading the virus. PSII, photosystem II.

are pathogens of immunocompromised humans. A virus that infects *Acanthamoeba polyphaga* was isolated and sequenced five years ago[32]. Mimivirus, as it was called, has a very large virion (more than 700 nm) and a genome of 1.2 megabases (Mb)[33], so it is larger and more genetically complex then many cellular organisms. Its large genome may be simply a by-product of having a large capsid (the genome's protein shell), in which case most of the genes may not be important for viral reproduction. If this is the case, these viruses serve as gigantic gene reservoirs. Evidence in favour of this hypothesis comes from the fact that the mimivirus genome is highly chimaeric and contains many genes related to the host[34]. However, nucleotide-composition studies suggest that horizontal gene transfer from the host is less common in large eukaryotic viruses than in phage[10].

A closely related strain of mimivirus, called mamavirus, adds another twist to the viral manipulation story[35]. Mamavirus (Fig. 2a) is infected by a satellite-phage-like entity, or 'virophage', called Sputnik (Fig. 2b). This 'virus of a virus' has an 18-kb genome. Inoculation of the host with both mamavirus and Sputnik increases the production of Sputnik and negatively affects mamavirus production. This is reminiscent of coliphages, in which P4 parasitizes the larger P2 phage. Recent mining of marine microbiomes has shown that viruses similar to large eukaryotic viruses are common and widely distributed in marine ecosystems[36,37].

Viral manipulation of protists

Coccolithophores are an abundant group of eukaryotic phytoplankton that are characterized by their intricate calcium carbonate scales, which are known as coccoliths (Fig. 2c). Blooms of coccolithophores influence global temperatures by increasing Earth's albedo (that is, more sunlight is reflected). Additionally, the sinking of the coccoliths and associated organic matter is one of the main mechanisms by which the ocean's biological pump draws down atmospheric carbon dioxide[38]. *Emiliania huxleyi*, named after Charles Darwin's advocate Thomas Huxley, is the most abundant species of coccolithophorid. It undergoes massive blooms that turn the sea a milky blue that is observable from satellites, but these blooms rapidly disappear. The main mechanism for these boom-and-bust cycles was thought to be infection and lysis by *E. huxleyi*-specific viruses. This hypothesis was first presented when large viral-like particles were found to co-occur with *E. huxleyi* and other phytoplankton in nutrient-augmented mesocosm experiments[39]. To test this hypothesis, viruses and microorganisms were sampled off the coast of Plymouth, UK, during a coccolithophore bloom[40]. Satellite images showed decreases in the light reflected from the coccolithophores at one of the sites. This area had lower concentrations of *E. huxleyi* cells but higher concentrations of viruses and free coccoliths. These data suggested that a viral lysis event had blown the coccolithophores apart. In support of this conclusion, two coccolithoviruses, EhV84 and EhV86, were isolated and sequenced from this bloom[41]. When the genome of EhV86 was sequenced, it turned out to be one of the largest known marine viral genomes (about 400 kb)[42].

The EhV86 genome is also notable for the presence of genes similar to those involved in ceramide production. Ceramides are involved in apoptosis (programmed cell death) and cell-cycle arrest. Ceramide production initiates apoptosis through the activation of caspases, the proteases that sit at the centre of the apoptotic pathway. Inhibiting the activity of metacaspase (a protein similar to caspases) in *E. huxleyi* effectively stops EhV1 production[43]. Furthermore, bioinformatic analysis has shown that many of

Table 1 | Some virally encoded proteins thought to modify the phenotypes of their marine hosts

Gene/Protein	Function	Host genus	Virus/Virome	Reference
PsbA	Photosynthesis	*Prochlorococcus* *Synechococcus*	P-SSP7, P-SSM2, P-SSM4 S-PM2	15 13
PsbD	Photosynthesis	*Prochlorococcus*	P-SSM4	17
Hli	Protection from photo-inhibition	*Prochlorococcus*	P-SSP7, P-SSM2, P-SSM4	14
PetE	Photosynthesis	*Prochlorococcus*	P-SSM2	14
PetF	Photosynthesis	*Prochlorococcus*	P-SSM2	14
TalC	Carbon metabolism	*Prochlorococcus*	P-SSM2	14
PstS	Phosphate recycling	*Roseobacter*	Roseophage SIO1	57
PhoH	Phosphate recycling	*Roseobacter*	Roseophage SIO1	57
Ceramide	Apoptosis	*Emiliania*	EhV86	43
CTX	Pathogenesis	*Vibrio*	CTXΦ	20

the viral proteins have caspase recognition sites, and cleavage of these sites by caspases is presumably necessary for viral production[44,45]. Although speculative, this viral manipulation of the host is interesting because the apoptotic pathway was probably once a mechanism to prevent the spread of viruses in coccolithophore blooms, but the pathway has been subverted to increase viral spread. As noted by the authors, this is a great example of the 'Red Queen' hypothesis in action[43]. As *E. huxleyi* develops resistance, the virus finds a way around it. They are in an arms race, where they need to 'run' (evolve) just to maintain their position (Fig. 1).

A wonderful twist to the tale of the coccolithophore and its virus comes from the study of an alternative escape route from the viral predators[46]. There are two distinct life stages in *E. huxleyi*[47]: the first is the diploid coccolith-bearing cell, which is the form most commonly studied; the second, haploid, stage exists as naked cells that can be motile or non-motile. It turns out that the haploid sexual stage of *E. huxleyi* is resistant to the viruses isolated from diploid cells[46], providing a way of avoiding viral infection and colony collapse. On the basis of these results, the authors proposed that the Red Queen hypothesis should be supplemented with the 'Cheshire Cat' model, in which there is a disappearing act rather than a running in place (Fig. 1).

Viral manipulation of metazoans

Marine viruses also infect and manipulate their metazoan hosts. One interesting example is the solar-powered sea slug[48]. These molluscs are some of the most fascinating creatures in the world, eating algae and harvesting chloroplasts through specialized epithelial cells in the gut (Fig. 3). Once phagocytosed, the chloroplasts are maintained within the animal's cells for months at a time, during which the slug gains energy directly from photosynthesis[49]. This process, which is effectively theft of the chloroplasts, is called kleptoplasty. However, plastid genomes only encode 10–20% of the genes necessary for photosynthesis, so where do the rest of the proteins for chloroplast function come from? To address this question, chloramphenicol was used to block protein synthesis by the chloroplast, and cycloheximide was used to inhibit eukaryotic ribosomes[50,51]. These experiments showed that some nuclear-encoded chloroplast proteins are synthesized by the slug cells and suggested that horizontal gene transfer occurred between the algal nuclear genome and the slug genome (plant-to-animal gene transfer)[52]. Recent genomic data show that this is indeed the case, and the most likely mechanism for that transfer is by way of a eukaryotic virus[53].

The best studied solar-powered slug is *Elysia chlorotica*, which eats only the alga *Vaucheria litorea* (Fig. 3c). The chloroplasts are absorbed by the slug's gut and are maintained for many months until the slugs lay eggs and suddenly die. This synchronous death is tightly correlated with the appearance of viral particles[53,54] (Fig. 3a, b), suggesting that the slug's annual life cycle is brought about by endogenous viruses. Supporting this hypothesis are the observations that there are no records of viral infection in juveniles and that animals maintained for months in the absence of food and other slugs still undergo this synchronous death.

Viral particles and crystalline arrays (Fig. 3b) have also been identified in the stolen chloroplasts and in the nuclei and cytoplasm of the host slug (Fig. 3a). Although the identity of these viruses remains unknown, reverse transcriptase activity has been detected during viral production stages, suggesting that vertically transmitted retroviruses are partly responsible for the deaths. The presence of viruses in both the chloroplast and the nucleus provides a hypothetical mechanism for the horizontal gene transfer of photosynthetic genes to the host[53]. Viruses, then, have two unexpected roles: first, they dramatically alter the slug's

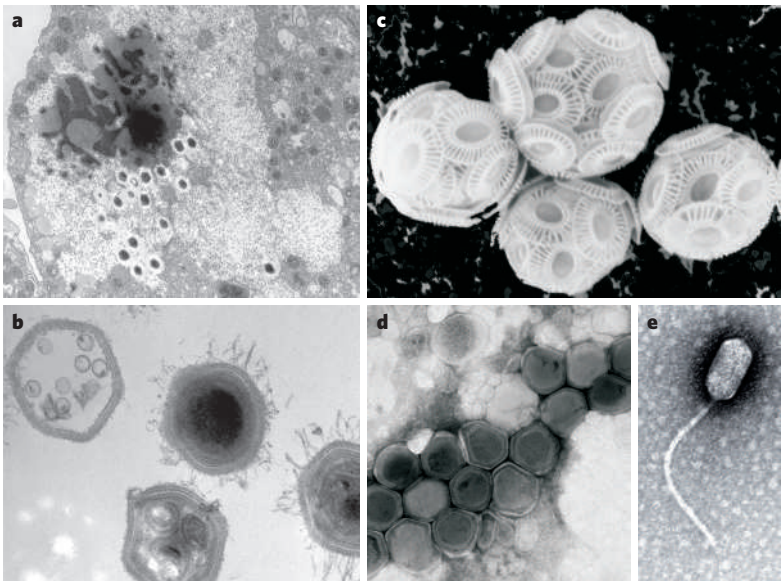


Figure 2 | Library of recently discovered marine viruses. Electron micrographs of mamavirus (**a**) and the virophage Sputnik (**b**). Mamavirus is a large icosahedral virus with a 1.2-Mb genome that infects the protist *Acanthamoeba*. Sputnik is a virus that infects mamavirus and lowers its fitness. *Emiliania huxleyi* (**c**), a coccolithophore important for marine primary production and nutrient recycling, and the *E. huxleyi*-like virus that causes boom-and-bust cycles and alters the life stages of its host (**d**). Phage of the cosmopolitan cyanobacterium *Prochlorococcus* (**e**). (Panels **a** and **b** courtesy of D. Raoult, Centre National de la Recherche Scientifique, Marseille, France. Panels **c** and **d** courtesy of W. Wilson, Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine. Panel **e** reproduced, with permission, from ref. 58.)

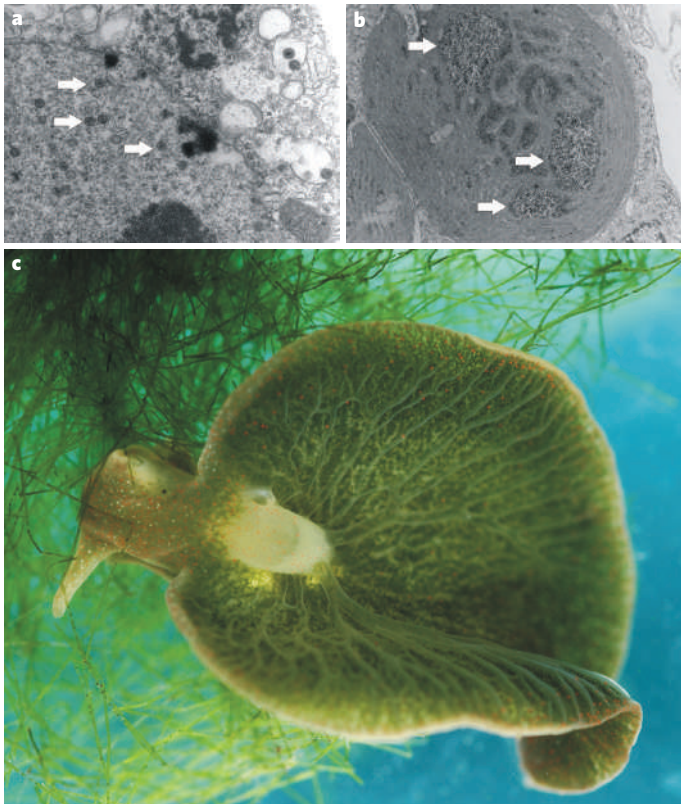


Figure 3 | Viruses that alter the life cycle of solar-powered slugs. Transmission electron micrograph of viral-like particles (arrows) in the nuclei and the cytoplasm (**a**) and within the stolen and incorporated algal chloroplasts (**b**) of the sea slug *Elysia chlorotica* (**c**), which uses solar energy. (Images courtesy of S. Piece, University of South Florida, Tampa.)

life history; and second, they are probably the vector for this horizontal gene transfer between an animal and a plant.

The future of marine virology

So far, the study of marine viruses has been dominated by the search for pathogens, but this will need to change if we are to appreciate the diverse ways that viruses affect life on Earth. We hope that future marine viral work will focus on three major areas of research. First, there needs to be a push to discover and evaluate viruses that infect marine archaeal organisms. Studies of these archaeal viruses — especially in the context of archaeal species that produce or consume greenhouse gases such as methane, or that help recycle limiting nutrients such as inorganic iron and nitrogen in the ocean — will yield interesting and important results. Second, we need to study the effects of viruses on the structure and function of zooplankton communities. Only a few studies have investigated the direct or indirect roles of viruses on zooplankton in the oceans. Culture studies have shown that viruses can contribute to boom-and-bust cycles in many metazoans, but no one has explored how the top-down forcing of metazoans by viral infections will affect zooplankton communities. Finally, recent work in the field of entomology has revealed remarkable tripartite symbioses between insects, bacteria and phage[55,56]. In these systems, the abundance of bacterial symbionts is controlled by viruses. This, in effect, changes the physiology and ecology of the host. These tripartite symbioses between eukaryote, microorganism and phage exemplify the intricacies of viral ecology that have so far been overlooked. We expect to find similarly complex interactions between marine organisms, their symbionts and the viruses that affect one or both. There is much exciting biology waiting to be discovered. ■

1. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5,** 801–812 (2007).
2. Brussaard, C. P. D. *et al.* Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J.* **2,** 575–578 (2008).
3. Steward, G. F., Montiel, J. L. & Azam, F. Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol. Oceanogr.* **45,** 1697–1706 (2000).
4. Wommack, K. E., Ravel, J., Hill, R. T., Chun, J. S. & Colwell, R. R. Population dynamics of Chesapeake bay virioplankton: total-community analysis by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* **65,** 231–240 (1999).
5. Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99,** 14250–14255 (2002).
6. Breitbart, M. *et al.* Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B* **271,** 565–574 (2004).
7. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4,** e368 (2006).
8. Desnues, C. *et al.* Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452,** 340–345 (2008).
9. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452,** 629–632 (2008).
This paper demonstrates that viromes contain many unexpected host genes.
10. Monier, A., Claverie, J.-M. & Ogata, H. Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* **8,** 456 (2007).
11. Paul, J. H. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* **2,** 579–589 (2008).
12. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63,** 106–127 (1999).
13. Mann, N. H. *et al.* The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* **187,** 3188–3200 (2005).
This paper shows that cyanophage genomes carry genes involved in photosynthesis.
14. Sullivan, M. B. *et al.* Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4,** e234 (2006).
15. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438,** 86–89 (2005).
16. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393,** 464–467 (1998).
17. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3,** e144 (2005).
18. Coleman, M. L. *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311,** 1768–1770 (2006).
19. Sharon, I. *et al.* Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* **1,** 492–501 (2007).
20. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272,** 1910–1914 (1996).
21. Williamson, S. J. *et al.* Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J.* **2,** 1112–1121 (2008).
22. Jiang, S. C. & Paul, J. H. Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* **64,** 2780–2787 (1998).
This pioneering study was the first to measure transduction rates in the marine environment.
23. Stanton, T. B. Prophage-like gene transfer agents — novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* **13,** 43–49 (2007).
24. Lang, A. S. & Beatty, J. T. Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* **15,** 54–62 (2007).
25. Biers, E. J. *et al.* Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl. Environ. Microbiol.* **74,** 2933–2939 (2008).
References 24 and 25 discuss the prevalence of GTAs and their role in horizontal gene transfer.
26. Chiura, H. X. Generalized gene transfer by virus-like particles from marine bacteria. *Aquat. Microb. Ecol.* **13,** 75–83 (1997).
27. Chiura, H. X. Broad host range xenotrophic gene transfer by virus-like particles from a hot spring. *Microbes Environ.* **17,** 53–58 (2001).
28. Breitbart, M., Miyake, J. H. & Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236,** 249–256 (2004).
29. Short, C. M. & Suttle, C. A. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* **71,** 480–486 (2005).
30. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. Movement of viruses between biomes. *Appl. Environ. Microbiol.* **70,** 5842–5846 (2004).
31. Visvesvara, G. S., Moura, H. & Schuster, F. L. Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris*, *Naegleria fowleri*, and *Sappinia diploidea*. *FEMS Immunol. Med. Microbiol.* **50,** 1–26 (2007).
32. La Scola, B. *et al.* A giant virus in amoebae. *Science* **299,** 2033 (2003).
33. Raoult, D. *et al.* The 1.2-megabase genome sequence of mimivirus. *Science* **306,** 1344–1350 (2004).
34. Filee, J., Siguier, P. & Chandler, M. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet.* **23,** 10–15 (2007).
35. La Scola, B. *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature* **455,** 100–104 (2008).
A parasite of mimiviruses, Sputnik phage, was identified and its genome characterized in this paper.
36. Monier, A., Claverie, J.-M. & Ogata, H. Taxonomic distribution of large DNA viruses in the sea. *Genome Biol.* **9,** R106 (2008).
37. Monier, A. *et al.* Marine mimivirus relatives are probably large algal viruses. *Virol. J.* **5,** 12 (2008).
38. Paasche, E. A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification–photosynthesis interactions. *Phycologia* **40,** 503–529 (2001).
39. Bratbak, G., Egge, J. K. & Heldal, M. Viral mortality of the marine alga *Emiliania huxleyi*

(Haptophyceae) and termination of algal blooms. *Mar. Ecol. Prog. Ser.* **93,** 39–48 (1993).
40. Wilson, W. H., Tarran, G. & Zubkov, M. V. Virus dynamics in a coccolithophore-dominated bloom in the North Sea. *Deep-Sea Res. II* **49,** 2951–2963 (2002).
41. Wilson, W. H. *et al.* Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J. Mar. Biol. Assoc. UK* **82,** 369–377 (2002).
42. Wilson, W. H. *et al.* Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* **309,** 1090–1092 (2005).
43. Bidle, K. D., Haramaty, L., Ramos, J. B. E. & Falkowski, P. Viral activation and recruitment of metacaspases in the unicellular coccolithophore, *Emiliania huxleyi*. *Proc. Natl Acad. Sci. USA* **104,** 6049–6054 (2007).
44. Best, S. M., Wolfinbarger, J. B. & Bloom, M. E. Caspase activation is required for permissive replication of Aleutian mink disease parvovirus *in vitro*. *Virology* **292,** 224–234 (2002).
45. Best, S. M. & Bloom, M. E. Caspase activation during virus infection: more than just the kiss of death? *Virology* **320,** 191–194 (2004).
46. Frada, M., Probert, I., Allen, M. J., Wilson, W. H. & de Vargas, C. The 'Cheshire Cat' escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proc. Natl Acad. Sci. USA* **105,** 15944–15949 (2008).
This paper shows that the different life stages of *E. huxleyi* are a way of avoiding viral mortality.
47. Klaveness, D. & Paasche, E. Two different *Coccolithus huxleyi* cell types incapable of coccolith formation. *Arch. Microbiol.* **75,** 382–385 (1971).
48. Mujer, C. V., Andrews, D. L., Manhart, J. R., Pierce, S. K. & Rumpho, M. E. Chloroplast genes are expressed during intracellular symbiotic association of *Vaucheria litorea* plastids with the sea slug *Elysia chlorotica*. *Proc. Natl Acad. Sci. USA* **93,** 12333–12338 (1996).
49. Green, B. J. *et al.* Mollusc–algal chloroplast endosymbiosis. Photosynthesis, thylakoid protein maintenance, and chloroplast gene expression continue for many months in the absence of the algal nucleus. *Plant Physiol.* **124,** 331–342 (2000).
50. Pierce, S. K., Curtis, N. E., Hanten, J. J., Boerner, S. L. & Schwartz, J. A. Transfer, integration and expression of functional nuclear genes between multicellular species. *Symbiosis* **43,** 57–64 (2007).
51. Pierce, S. K., Curtis, N. E., Schwartz, J. A. & Massey, S. E. Functional algal nuclear genes are present in a sea slug genome — horizontal gene transfer demonstrated. *Integr. Comp. Biol.* **46,** e110 (2006).
52. Pierce, S. K., Massey, S. E., Hanten, J. J. & Curtis, N. E. Horizontal transfer of functional nuclear genes between multicellular organisms. *Biol. Bull.* **204,** 237–240 (2003).
53. Pierce, S. K., Maugel, T. K., Rumpho, M. E., Hanten, J. J. & Mondy, W. L. Annual viral expression in a sea slug population: Life cycle control and symbiotic chloroplast maintenance. *Biol. Bull.* **197,** 1–6 (1999).
In this study, viruses are implicated as a route to the horizontal transfer of photosynthetic genes from chloroplasts to metazoan host genomes.
54. Mondy, W. L. & Pierce, S. K. Apoptotic-like morphology is associated with annual synchronized death in kleptoplastic sea slugs (*Elysia chlorotica*). *Invertebr. Biol.* **122,** 126–137 (2003).
55. Bordenstein, S. R., Marshall, M. L., Fry, A. J., Kim, U. & Wernegreen, J. J. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog.* **2,** e43 (2006).
56. Moran, N. A., Degnan, P. H., Santos, S. R., Dunbar, H. E. & Ochman, H. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc. Natl Acad. Sci. USA* **102,** 16919–16926 (2005).
57. Mannisto, R. H., Kivela, H. M., Paulin, L., Bamford, D. H. & Bamford, J. K. H. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. *Virology* **262,** 355–363 (1999).
58. Fuhrman, J. Genome sequences from the sea. *Nature* **424,** 1001–1002 (2003).
59. Zobell, C. E. *Marine Microbiology* (Chronica Botanica, 1946).
60. Spencer, R. A marine bacteriophage. *Nature* **175,** 690–691 (1955).
61. Pomeroy, L. R. The ocean's food web, a changing paradigm. *Bioscience* **24,** 499–504 (1974).
62. Torrella, F. & Morita, R. Y. Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon: ecological and taxonomical implications. *Appl. Environ. Microbiol.* **37,** 774–778 (1979).
63. Moebus, K. A method for the detection of bacteriophages from ocean water. *Helgol. Meeresunters.* **34,** 1–14 (1980).
64. Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40,** 1236–1242 (1995).
65. Suttle, C. A. The significance of viruses to mortality in aquatic microbial communities. *Microb. Ecol.* **28,** 237–243 (1994).
66. Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M. & Sanudo-Wilhelmy, S. A. Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. *Limnol. Oceanogr.* **42,** 1492–1504 (1997).
67. Rohwer, F. *et al.* The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol. Oceanogr.* **45,** 408–418 (2000).
68. Culley, A. I., Lang, A. S. & Suttle, C. A. Metagenomic analysis of coastal RNA virus communities. *Science* **312,** 1795–1798 (2006).
69. Culley, A. I., Lang, A. S. & Suttle, C. A. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virol. J.* **4,** 69 (2007).
70. Comeau, A. M., Chan, A. M. & Suttle, C. A. Genetic richness of vibriophages isolated in a coastal environment. *Environ. Microbiol.* **8,** 1164–1176 (2006).
71. Jiang, S. C. & Paul, J. H. Significance of lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. *Microb. Ecol.* **35,** 235–243 (1998).
72. Paul, J. H. *et al.* Complete genome sequence of Φ HSIC, a pseudotemperate marine phage of *Listonella pelagia*. *Appl. Environ. Microbiol.* **71,** 3311–3320 (2005).
73. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424,** 741 (2003).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to R.V.T. (rvegathurber@gmail.com).

ARTICLES

A surface transporter family conveys the trypanosome differentiation signal

Samuel Dean[1], Rosa Marchetti[2], Kiaran Kirk[2] & Keith R. Matthews[1]

Microbial pathogens use environmental cues to trigger the developmental events needed to infect mammalian hosts or transmit to disease vectors. The parasites causing African sleeping sickness respond to citrate or *cis*-aconitate (CCA) to initiate life-cycle development when transmitted to their tsetse fly vector. This requires hypersensitization of the parasites to CCA by exposure to low temperature, conditions encountered after tsetse fly feeding at dusk or dawn. Here we identify a carboxylate-transporter family, PAD (proteins associated with differentiation), required for perception of this differentiation signal. Consistent with predictions for the response of trypanosomes to CCA, PAD proteins are expressed on the surface of the transmission-competent 'stumpy-form' parasites in the bloodstream, and at least one member is thermoregulated, showing elevated expression and surface access at low temperature. Moreover, RNA-interference-mediated ablation of PAD expression diminishes CCA-induced differentiation and eliminates CCA hypersensitivity under cold-shock conditions. As well as being molecular transducers of the differentiation signal in these parasites, PAD proteins provide the first example of a surface marker able to discriminate the transmission stage of trypanosomes in their mammalian host.

Insect-borne parasites undergo life-cycle differentiation to adapt to rapid changes in temperature[1,2], nutritional availability[3] and potential immunological attack[4] as they enter their arthropod vector. The cues that induce such changes are often well characterized, such as temperature reduction and pH changes[5] or exposure to arthropod-derived factors[6]. However, the surface molecules that transmit these signals and initiate intracellular differentiation events in microbial parasites are often not well characterized.

African trypanosomes are protozoan parasites responsible for fatal disease in humans and livestock in sub-Saharan Africa, generating significant restrictions in health and welfare in afflicted regions[7]. The transmission of these parasites by tsetse flies requires the development of bloodstream 'stumpy forms', a G0-arrested cell-type pre-adapted for transmission[8–10]. Stumpy forms arise via quorum sensing from proliferative 'slender forms' at the peak of each parasitaemia in response to an unidentified parasite-derived signalling factor, stumpy induction factor (SIF)[11]. Although the morphological extremes of slender and stumpy forms are easily distinguished, the development from slender to stumpy morphology is progressive[12], with transitional forms loosely described as 'intermediates'. So far, no developmentally regulated surface marker has been identified that discriminates between slender and stumpy cells.

When ingested by tsetse flies during a blood meal, slender forms are killed whereas stumpy forms differentiate to procyclic forms[13]. Bloodstream trypanosomes can also be induced to differentiate *in vitro* by the Krebs-cycle intermediates citrate or *cis*-aconitate (CCA)[14]. Recently, it was discovered that temperature reduction from 37 °C to 20 °C could induce hypersensitivity of stumpy forms to CCA[2], such that differentiation was induced at concentrations found in the tsetse fly (15.9 μM[15]) or ingested blood meal (~130 μM[16]). This probably represents a natural condition in the trypanosome life cycle, as tsetse flies are exposed to cool conditions when feeding at dusk or early morning[2]. Nonetheless, the molecule responsible for the transmission of the CCA differentiation signal has remained unidentified.

PAD family identification and expression

We previously selected a trypanosome line (DiD1, defective in differentiation-1) with reduced ability to differentiate to procyclic forms[10]. The expression profile of DiD1 was compared with its differentiation-competent parent by differential hybridization of labelled cDNA to genomic arrays. This identified two adjacent genes on chromosome 7 of the *Trypanosoma brucei* genome (Tb927.7.5930 and Tb927.7.5940), there being a stronger signal with DiD1-derived cDNA compared to the parental. These genes comprise the first two genes in an eight-gene array at the end of a unidirectional gene cluster (named the '*PAD*' gene array, for 'proteins associated with differentiation'; see below) (Supplementary Fig. 1a). Although no mutation in either gene was detected (Supplementary Fig. 2), northern blotting confirmed the differential expression of both *PAD1* and *PAD2* genes in the DiD1 line (Fig. 1a). Furthermore, *PAD1* and *PAD2* messenger RNA showed stage-regulated expression, *PAD1* being enriched in stumpy forms, whereas *PAD2* was elevated in procyclic forms. Neither *PAD1* nor *PAD2* mRNA was significantly expressed in slender forms (Fig. 1a).

The *PAD* genes encode closely related members of a family of 14 transmembrane-spanning proteins of the major facilitator superfamily (Fig. 1b; see also Supplementary Fig. 1b). PSI-BLAST searches revealed a conserved domain in plant nodulin-like proteins (PF06813) and closest overall similarity to carboxylate transporters. Supporting this, *Xenopus* oocytes microinjected with cRNA encoding either PAD1 or PAD2 showed a marked increase in the uptake of [^{14}C]citrate relative to non-injected controls (Fig. 1c).

An antibody detecting all members of the PAD protein family reacted with distinct bands in different trypanosome life-cycle stages (Fig. 1d); although there was no significant expression in slender forms, stumpy forms showed two prominent bands at 55 kDa and 57 kDa, whereas procyclic forms predominantly expressed the 57 kDa band. Neither band corresponded to the predicted size of any PAD protein, probably because proteins with extensive transmembrane regions

[1]Centre for Immunity, Infection and Evolution, Institute for Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. [2]School of Biochemistry & Molecular Biology, Australian National University, Canberra, ACT 0200, Australia.

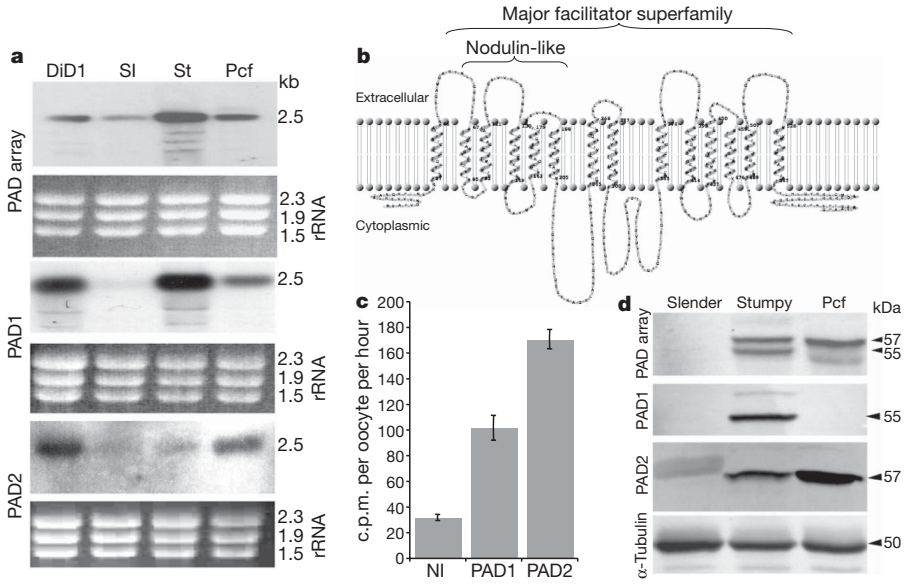


Figure 1 | Identification and characteristics of PAD proteins. **a**, Expression of *PAD* mRNAs in DiD1, parental slender forms (Sl), stumpy forms (St) and procyclic forms (Pcf). rRNAs show loading. **b**, Predicted transmembrane domains in PAD1. **c**, [^{14}C]citrate uptake into *Xenopus* oocytes microinjected with 20 ng cRNA encoding PAD1 or PAD2, or non-injected (NI). The data (from 8–10 oocytes, shown ± s.e.m.) are representative of ten experiments in which oocytes expressing PAD1 exhibited 1.4–3.7-fold increase in [^{14}C]citrate uptake and oocytes expressing PAD2 exhibited 1.6–5.4-fold increase relative to non-injected controls. c.p.m., counts per minute. **d**, Expression of PAD proteins in slender, stumpy and procyclic forms normalized to α-tubulin. The PAD2 antibody cross-reacts weakly with the 60 kDa VSG in slender forms.

frequently migrate aberrantly. Specific anti-peptide antibodies showed that the 55 kDa and 57 kDa bands corresponded to PAD1 and PAD2 respectively, and that PAD1 was stumpy-form-specific, whereas PAD2 was expressed sixfold more in procyclic forms than stumpy forms (Fig. 1d). During synchronous differentiation from stumpy to procyclic forms, PAD1 was retained only during the first 24 h, whereas PAD2 was strongly induced during this period (up to 17-fold at 24 h; Supplementary Fig. 3).

PAD1 marks the transmissible stumpy form

The PAD1 expression profile suggested that it might provide a useful cytological marker for stumpy forms. To investigate this, the location and expression of PAD1 was analysed in mixed populations of slender and stumpy forms. Figure 2a (left and middle panels) shows immunofluorescence images demonstrating stumpy-specific expression of PAD1. Confocal images (Fig. 2a, right panel) demonstrated an intense staining at the stumpy-cell periphery, revealing surface membrane labelling. To quantify the stumpy-specific expression of PAD1, 'intermediate' cell populations were assayed with PAD1 and counterstained with 4,6-diamidino-2-phenylindole (DAPI), allowing their cell-cycle position to be determined. In trypanosomes, cells at the G1/G0 and S phase have one kinetoplast and one nucleus (1K1N), whereas G2/M and post-mitotic cells have a 2K1N or 2K2N DNA configuration, respectively[17]. Because stumpy forms are uniformly arrested at G1/G0 (ref. 18), 2K1N or 2K2N cells can be unambiguously assigned as slender forms. Detailed examination of cells within each category of 1K1N, 2K1N or 2K2N (an overall analysis of >10,000 cells) demonstrated that dividing cells and cells with a slender morphology were overwhelmingly negative for PAD1 (Fig. 2b). However, ~10% of 1K1N and 2K2N slender cells were PAD1-positive—these may represent intermediate cells that have committed to stumpy formation.

To investigate whether only the PAD1-positive bloodstream trypanosomes were competent to differentiate to procyclic forms, mixed populations of slender and stumpy forms were examined 6 h

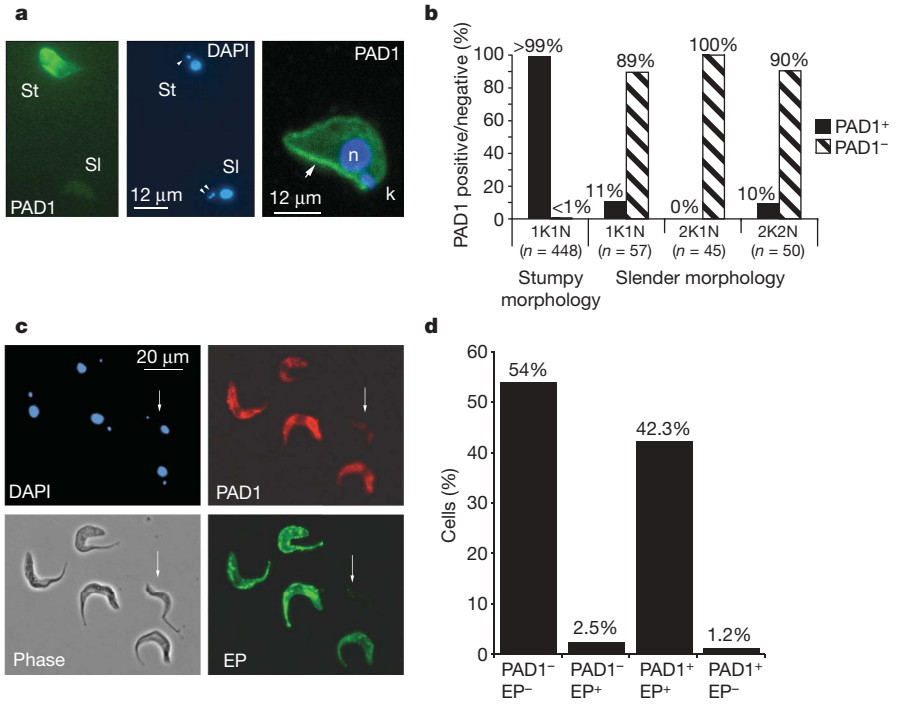


Figure 2 | PAD1 identifies stumpy forms. **a**, PAD1 expression on methanol-fixed trypanosomes. The slender cell (sl; identified by having two kinetoplasts, indicated by arrowheads in the DAPI panel) is PAD1 negative, the stumpy cell (St) is PAD1 positive. A confocal image of a paraformaldehyde-fixed stumpy form is also shown (right); peripheral PAD1 labelling (green) is indicated with an arrow. n, nucleus; k, kinetoplast (both stained blue). **b**, Quantification of PAD1 expression on different cell-cycle stages. **c**, PAD1-positive cells are those competent to differentiate. Slender and stumpy forms were methanol-fixed 6 h through differentiation. A slender-morphology PAD1-negative, EP-procyclin-negative cell is indicated by an arrow; other cells are PAD1-positive (red), EP-procyclin-positive (green). Phase-contrast and DAPI images are also shown. **d**, Quantification of the cells expressing PAD1 and/or EP-procyclin. $n = 500$.

after exposure to *cis*-aconitate (CA) for EP-procyclin expression, an early marker for differentiation[19]. This confirmed a precise correspondence between EP-procyclin and PAD1 expression, with 96% of cells showing matching staining for both markers (Fig. 2c, d). Thus, PAD1 identifies stumpy forms in a bloodstream population as those competent to differentiate to procyclic forms.

PAD2 is thermoregulated

As surface carboxylate transporters expressed on stumpy forms, PAD proteins were good candidates as transducers of the CCA developmental signal. To evaluate this, we initially tested the inducibility of PAD proteins at 20 °C, a predicted requirement for physiological CCA sensitivity[2]. This revealed that PAD2 was consistently upregulated ~4-fold (range 3–32-fold) at 20 °C. Analysing two trypanosome strains (*T. brucei* AnTat1.1 and *T. brucei* EATRO 2340) demonstrated fourfold and 3.8-fold upregulation of PAD2, respectively, at 20 °C when compared to 37 °C (Fig. 3a). In contrast, PAD1 expression was unaltered. This eliminated the possibility that the elevation of PAD2 expression was due to general enhancement of membrane protein expression at low temperature. Other stress conditions such as pH or cell concentration did not affect PAD2 expression (Supplementary Fig. 4 and data not shown). Notably, at 37 °C PAD2 was predominantly at the flagellar pocket, whereas it was surface located at 20 °C (Fig. 3b, c), the redistribution occurring within 60 min (Supplementary Fig. 5). Thus, PAD2 is a novel example of both a thermoregulated parasite molecule[1] and a transmembrane protein showing regulated surface distribution[20].

PAD proteins convey the CCA differentiation signal

To functionally test PAD proteins in CCA-initiated differentiation, a gene fragment with >95% identity between each *PAD* gene was targeted by RNA interference (RNAi), enabling simultaneous knockdown of all members. This was performed in the pleomorphic *T. brucei* AnTat1.1 90:13 line[2] that generates stumpy forms at high frequency, exhibits cold sensitivity to CCA and has been engineered for doxycycline-inducible RNAi-mediated transcript ablation. The resulting transgenic cell line, in parallel with the parental line, was grown in mice with or without doxycycline induction for 6 days, each producing >80% stumpy populations (Supplementary Fig. 6). Although the RNAi effect was considerably leaky, there was 80% depletion of PAD proteins in the RNAi-induced cells (Fig. 4a), demonstrating that significant PAD expression was not required for either stumpy formation or survival (Supplementary Fig. 6). Once the transgenic stumpy forms were harvested from blood and incubated for 16 h at 20 °C, they were exposed to different concentrations of CA and differentiation monitored by EP-procyclin expression. Cells incubated at 20 °C showed the expected cold induction of EP-procyclin expression[2], although this was consistently and inducibly reduced in the RNAi line (Fig. 4b, '0 h CS' samples). This may indicate some interaction between PAD and EP-procyclin surface expression similar to mammalian monocarboxylate transporters, which co-associate with single-pass membrane proteins for surface access and activity[21]. Notably, PAD depletion reduced the differentiation of cells exposed to CA, with this being more effective at 20 °C (general linear model, $F_{1,44} = 16.82$, $P < 0.0005$ at 6 h) than 37 °C ($F_{1,44} = 6.60$, $P < 0.014$). Indeed, at 0.1 mM CA there was a 70% reduction of EP-procyclin expression compared to the control line at 6 h, with this expression being further reduced over 24 h due to reversal of cold induction at 27 °C (Fig. 4c; see also Supplementary Fig. 7). Matching the response to CA, *PAD* RNAi also diminished the response of stumpy cells to citrate at 20 °C (Supplementary Fig. 8). A cell-line-specific differentiation defect unrelated to *PAD* RNAi was eliminated by transiently transfecting pleomorphic cells with the RNAi construct; this recapitulated the differentiation phenotype (Supplementary Fig. 9). Moreover, the RNAi lines differentiated as well as parental cells in response to pronase treatment, an alternative differentiation trigger (Supplementary Fig. 10)[15]. Hence, ablating

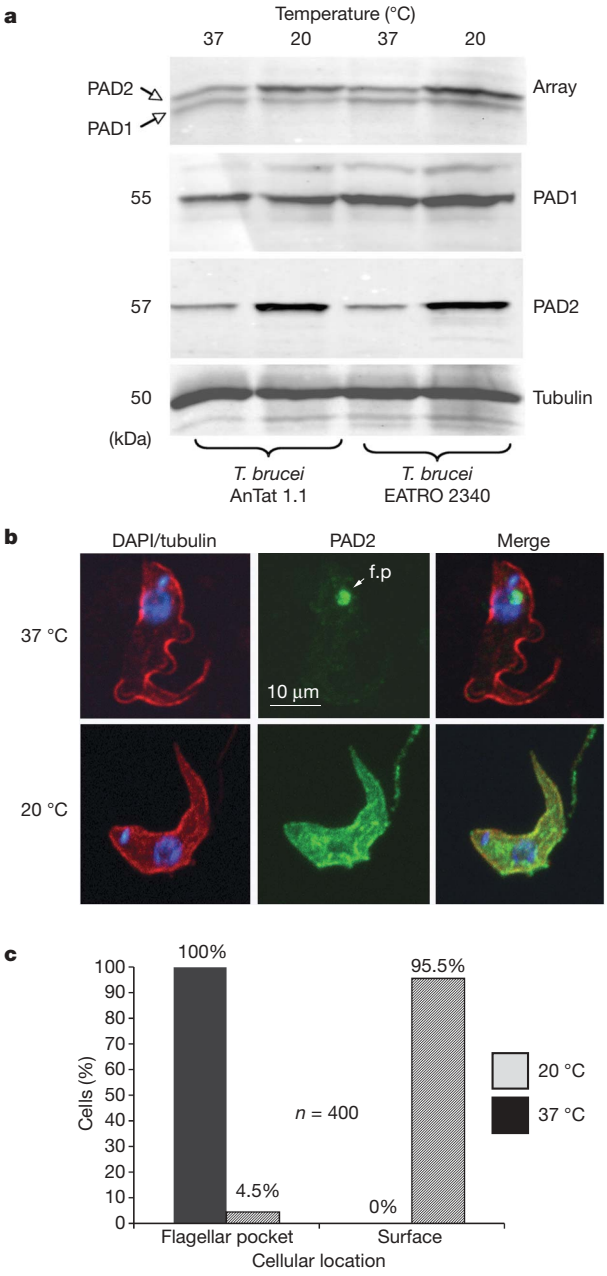


Figure 3 | PAD2 is cold-inducible. a, PAD protein expression at 37 °C or 20 °C in stumpy forms of *T. brucei* EATRO 2340 or *T. brucei* AnTat1.1. Samples were stained for all PAD proteins ('Array'), PAD1 or PAD2. α-Tubulin controlled for loading. **b**, Confocal immunofluorescence images of paraformaldehyde-fixed stumpy forms incubated at 37 °C or 20 °C and co-labelled for α-tubulin (red), PAD2 (green) and DAPI (blue). At 37 °C, PAD2 predominantly localized to the flagellar pocket (f.p.; indicated with an arrow); at 20 °C PAD2 located at the cell surface. **c**, Quantification of the PAD2 location at 37 °C or 20 °C.

PAD mRNAs reduced overall differentiation responses to CCA at all concentrations tested but specifically abrogated the cold-induced CCA hypersensitivity of stumpy forms.

Conclusions

These experiments demonstrate that the PAD proteins act as transducers of the CCA differentiation signal in *Trypanosoma brucei*. This conclusion is based on several lines of evidence: (1) PAD proteins are surface molecules expressed on stumpy forms but absent in transmission-incompetent slender forms; (2) at the single-cell level, PAD protein

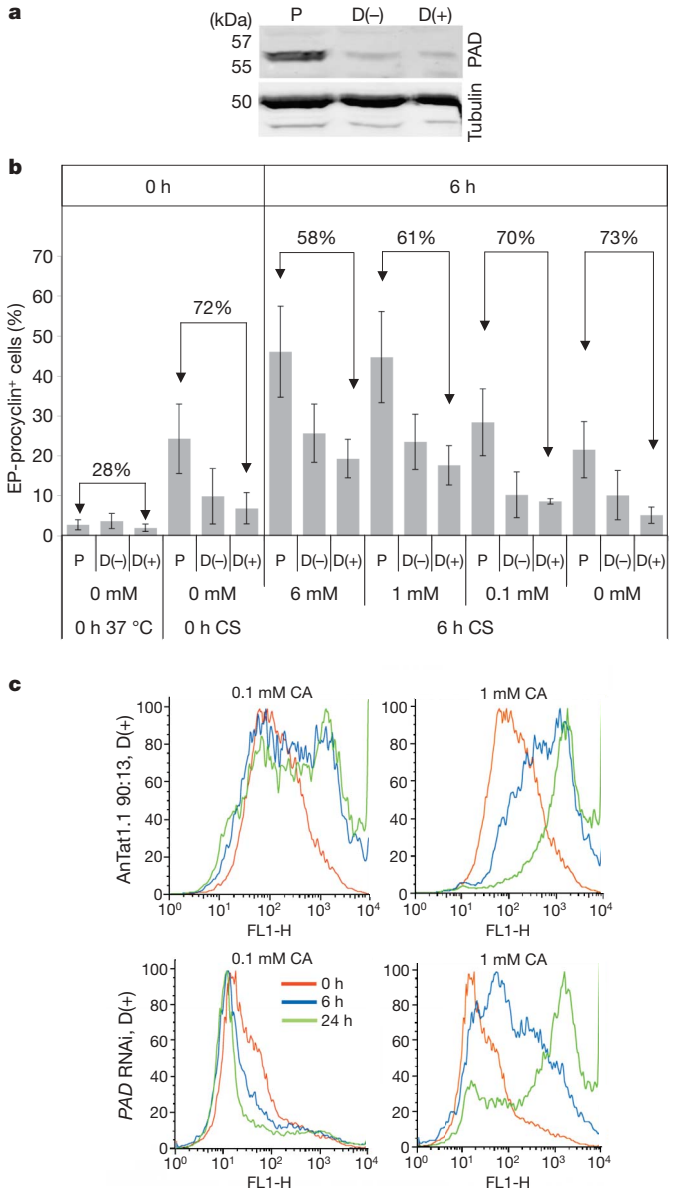


Figure 4 | RNAi against all *PAD* genes reduces differentiation. a, PAD-array expression in parental or *PAD* RNAi lines (with (D(+)) or without (D(−)) doxycycline) after 16 h at 20 °C. Tubulin, loading control. **b**, EP-procyclin in parental (P) and *PAD* RNAi stumpy forms grown with or without doxycycline, after 16 h at 20 °C, then incubated at 27 °C with a *cis*-aconitate (CA) titration. Means (±s.e.m.) of three experiments are shown, as is the percentage reduction in EP-procyclin in the *PAD* RNAi cells compared with parental cells. **c**, Flow cytometry of EP-procyclin expression in *PAD* RNAi or parental cells incubated with 0.1 mM or 1 mM CA. Cold-induced hypersensitivity to CA is ablated in the *PAD* RNAi line; at ≥1 mM CA both populations differentiate, although this is reduced in the RNAi line. Full flow cytometry data are available in Supplementary Fig. 7.

expression correlates precisely with the differentiation capacity of bloodstream-form parasites; (3) at least one PAD protein (PAD2) demonstrates cold-regulated expression and localization, consistent with predictions for the reception of the CCA signal *in vivo*; and (4) depletion of PAD protein expression in pleomorphic trypanosomes reduces their competence for differentiation and eliminates CCA responsiveness at physiologically relevant concentrations. In laboratory-adapted monomorphic slender trypanosomes, which do not express detectable levels of PAD proteins but which can differentiate in response to CCA, signalling is probably driven through other transporters by the high concentrations of CCA required for differentiation in these cell lines[14]. Consistent with this, incubating monomorphic slender forms, *T. brucei* AnTat1.1 90:13 stumpy forms and the *PAD* RNAi stumpy forms with [^{14}C]citrate revealed a clear correlation between PAD expression and cell-associated label (Supplementary Fig. 11). All of these characteristics are compatible with current models for the reception of the signal to differentiation from bloodstream to procyclic forms[8], with the relay of the CCA signal by PAD proteins providing the first molecular insight linking environmental sensing to trypanosome cell-type differentiation. Notably, this does not exclude an additional or complementary role of proteases in the tsetse fly midgut, which can act with CCA to promote robust differentiation[22].

The identification of PAD1 as a stumpy-specific surface marker protein also provides two key advances. First, the precise correlation between PAD1 expression and differentiation capacity directly confirms stumpy forms as the essential transmissible stage in the bloodstream, supporting classical observations[23] but contrasting with some recent models[24]. Second, the discovery of PAD1 permits the quantitative modelling of trypanosome population dynamics in chronic infections[25] and the development of bioassays to detect stumpy formation. This has obvious application for monitoring the activity of the signal for quorum sensing, SIF and in high-throughput screens for therapeutic agents that promote stumpy formation and hence prevent parasite virulence.

METHODS SUMMARY

***PAD* RNAi constructs.** PAD protein sequences from GeneDB were analysed using TMHMM[26] and displayed using TMRPres2D[27] to determine the position of the transmembrane helices. The *PAD1* reading frame was amplified from genomic DNA by polymerase chain reaction using 5′-TTTAAGCTTTGATCAA TGAGCGCACCCGTCGACAACGTC-3′ and 5′-AAACTCGAGCATATGTCAT TGCGGAGCAGCCTCACGGGC-3′ primers, and cloned into the HindIII-XbaI and XhoI-BamHI cloning sites of pALC14 (ref. 28) to generate the *PAD* RNAi plasmid.

Parasite growth and transfection. Culturing, transfection, differentiation and cold-shock assays were performed as described[2,29] on *T. brucei* AnTat1.1 90:13 (ref. 2) and selected using 0.5 μg ml^{-1} puromycin. Stumpy-enriched populations were obtained by DEAE-cellulose purification[30] of parasites 6–7 days after infection into cyclophosphamide-treated mice.

PAD expression analysis. Peptides specific to PAD1 (H_2N-CPKEPTRDAREAA PQ-COOH and H_2N-ETCCRREVAE-COOH), PAD2 (H_2N-EAEDNQTNAEN VC-COOH and H_2N-CNADACLEEKAADSSK-COOH) and the entire array (H_2N-VETDVDYIAPQFQET-CONH$_2$ and H_2N-TQQADKLGQDVCTER-COOH) were used to generate antipeptide antibodies (Eurogentec). Immunofluorescence analysis was performed[31] using a Zeiss Axioscope 2 or Leica SP5 confocal microscope and analysed using Volocity software (Improvision). Images were processed using Adobe Photoshop CS. Western blotting was performed by low-voltage SDS–PAGE and wet transfer onto Immobilon-P PVDF (Millipore) according to the manufacturer's instructions; proteins were detected using the LI-COR Odyssey system for quantification against a tubulin loading control. Flow cytometry analysis was performed using the FACS-Calibur flow cytometer (Becton Dickenson)[11].

***Xenopus* oocyte transport assays.** *PAD* genes were cloned into pGHJ and radiolabelled citrate uptake experiments were performed as described[32]. Uptake measurements were made over 1 h in oocytes (5 days after injection) incubated at pH 9 and 27.5 °C.

Statistical analysis. Differentiation data were analysed using a general linear model. Residuals did not conform to a normal distribution and therefore a logarithmic transformation was used. Statistical analysis was carried out using minitab version 15, with P values of $P < 0.05$ being considered statistically significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 February; accepted 17 March 2009.

1. Fang, J. & McCutchan, T. F. Thermoregulation in a parasite's life cycle. *Nature* **418**, 742 (2002).
2. Engstler, M. & Boshart, M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. *Genes Dev.* **18**, 2798–2811 (2004).
3. Lee, S. H., Stephens, J. L., Paul, K. S. & Englund, P. T. Fatty acid synthesis by elongases in trypanosomes. *Cell* **126**, 691–699 (2006).

4. Hao, Z. *et al.* Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc. Natl Acad. Sci. USA* **98,** 12648–12653 (2001).
5. Zilberstein, D. & Shapira, M. The role of pH and temperature in the development of *Leishmania* parasites. *Annu. Rev. Microbiol.* **48,** 449–470 (1994).
6. Billker, O. *et al.* Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* **392,** 289–292 (1998).
7. Barrett, M. P. *et al.* The trypanosomiases. *Lancet* **362,** 1469–1480 (2003).
8. Fenn, K. & Matthews, K. R. The cell biology of *Trypanosoma brucei* differentiation. *Curr. Opin. Microbiol.* **10,** 539–546 (2007).
9. Szoor, B. *et al.* Protein tyrosine phosphatase TbPTP1: a molecular switch controlling life cycle differentiation in trypanosomes. *J. Cell Biol.* **175,** 293–303 (2006).
10. Tasker, M. *et al.* A novel selection regime for differentiation defects demonstrates an essential role for the stumpy form in the life cycle of the African trypanosome. *Mol. Biol. Cell* **11,** 1905–1917 (2000).
11. Vassella, E., Reuner, B., Yutzy, B. & Boshart, M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J. Cell Sci.* **110,** 2661–2671 (1997).
12. Bruce, D. *et al.* The morphology of the trypanosome causing disease in man in Nyasaland. *Proc. R. Soc. Lond. B* **85,** 423–433 (1912).
13. Turner, C. M., Barry, J. D. & Vickerman, K. Loss of variable antigen during transformation of *Trypanosoma brucei rhodesiense* from bloodstream to procyclic forms in the tsetse fly. *Parasitol. Res.* **74,** 507–511 (1988).
14. Czichos, J., Nonnengaesser, C. & Overath, P. *Trypanosoma brucei*: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes *in vitro*. *Exp. Parasitol.* **62,** 283–291 (1986).
15. Hunt, M., Brun, R. & Kohler, P. Studies on compounds promoting the *in vitro* transformation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Parasitol. Res.* **80,** 600–606 (1994).
16. Jacobs, S. L. & Lee, N. D. Determination of citric acid in serum and urine using Br82. *J. Nucl. Med.* **5,** 297–301 (1964).
17. Sherwin, T. & Gull, K. The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Phil. Trans. Roy. Soc. Lond.* **323,** 573–588 (1989).
18. Shapiro, S. Z., Naessen, J., Liesegang, B., Moloo, S. K. & Magondu, J. Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Trop.* **41,** 313–323 (1984).
19. Matthews, K. R. & Gull, K. Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J. Cell Biol.* **125,** 1147–1156 (1994).
20. Kamsteeg, E. J. *et al.* MAL decreases the internalization of the aquaporin-2 water channel. *Proc. Natl Acad. Sci. USA* **104,** 16696–16701 (2007).
21. Kirk, P. *et al.* CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression. *EMBO J.* **19,** 3896–3904 (2000).
22. Sbicego, S. *et al.* The use of transgenic *Trypanosoma brucei* to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. *Mol. Biochem. Parasitol.* **104,** 311–322 (1999).
23. Robertson, M. Notes on the polymorphism of *Trypanosoma gambiense* in the blood and its relation to the exogenous cycle in *Glossina palpalis*. *Proc. R. Soc. Lond. B* **85,** 241–248 (1912).
24. Bass, K. E. & Wang, C. C. The *in vitro* differentiation of pleomorphic *Trypanosoma brucei* from bloodstream into procyclic form requires neither intermediary nor short-stumpy stage. *Mol. Biochem. Parasitol.* **44,** 261–270 (1991).
25. Lythgoe, K. A., Morrison, L. J., Read, A. F. & Barry, J. D. Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc. Natl Acad. Sci. USA* **104,** 8095–8100 (2007).
26. Tusnady, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17,** 849–850 (2001).
27. Spyropoulos, I. C., Liakopoulos, T. D., Bagos, P. G. & Hamodrakas, S. J. TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics* **20,** 3258–3260 (2004).
28. Pusnik, M. *et al.* Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol. Cell. Biol.* **27,** 6876–6888 (2007).
29. McCulloch, R. *et al.* Transformation of monomorphic and pleomorphic *Trypanosoma brucei*. *Methods Mol. Biol.* **262,** 53–86 (2004).
30. Lanham, S. M. Separation of trypanosomes from the blood of infected rats and mice by anion-exchangers. *Nature* **218,** 1273–1274 (1968).
31. Field, M. C. *et al.* New approaches to the microscopic imaging of *Trypanosoma brucei*. *Microsc. Microanal.* **10,** 621–636 (2004).
32. Saliba, K. J. *et al.* Sodium-dependent uptake of inorganic phosphate by the intracellular malaria parasite. *Nature* **443,** 582–585 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Ullu for the gift of genomic macroarrays, M. Engstler and M. Boshart for the gift of the AnTat1.1 90:13 line, and A. Schneider for pALC14. We thank D. Hall, P. Davies and D. Levin for technical assistance, P. MacGregor for statistical analysis and A. Paterou and D. Murray for Image analysis. This work was supported by a Wellcome Trust project grant and programme grant to K.R.M. S.D. was supported by a BBSRC studentship, a Wellcome Trust Programme Grant to K.R.M. and by a Journal of Cell Science Travelling fellowship for a visit to the laboratory of K.K. Support was also provided through a Wellcome Trust strategic award for the Centre for Immunity, Infection and Evolution and a BBSRC REI award for confocal facilities.

Author Contributions S.D. carried out all trypanosome experiments, R.M. carried out the *Xenopus* oocyte transport assays, K.K. contributed to the design of the transport assays, and K.R.M. conceived and supervised the study; the manuscript was written by K.R.M. and S.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.R.M. (keith.matthews@ed.ac.uk).

METHODS

Trypanosomes. *Trypanosoma brucei brucei* AnTat1.1 and *Trypanosoma brucei brucei* EATRO 2340 were used. Stumpy forms were generated by 5–6 days growth in MF1 mice, treated with cyclophosphamide 24 h before infection. Slender and monomorphic slender parasites were generated after 3 days growth in rodents.

For pleomorphic transfection, cells from an early-stage parasitaemia (3–4 days, depending on the infection) were purified from the buffy coat and the cells transfected using an AMAXA nucleofector protocol (T-cell nucleofection buffer, programme X001). Drug selection was carried out in cells maintained in HMI-9 media supplemented with 1.1% methyl cellulose at a concentration of approximately 5×10^6 cells ml^{-1}. Transgenic parasites were selected using 0.5 μg ml^{-1} puromycin.

Macroarray analysis. Expression differences between the DiD1 and parental progenitor were carried out by reverse transcription of 5 μg of poly(A)$^+$ RNA from each cell type, this being labelled using the GE Healthcare Gene Images non-radioactive labelling system. Macroarrays (a gift from E. Ullu) comprised 15,000 plasmid clones of 1–2 kb sheared DNA (~0.8× genome coverage) from *T. brucei* TREU 927/4 cloned into pUC18 and arrayed in 384 microtitre dishes over 39 plates. Bacterial clones were spotted onto 22 cm^2 nylon membranes in a 4×4 array format. After hybridization, signals were detected by ECL and differential signals, obtained in duplicate, were identified and validated by northern analysis.

Western blot analysis of PAD proteins. Cells were re-suspended at 3×10^6 cells in 10 μl of Laemmli sample buffer containing β-mercaptoethanol at room temperature. Genomic DNA was then sheared by sonication and the sample placed on ice. A total of 10 μl of protein sample was then resolved on a 10% polyacrylamide gel at 100 V at 4 °C for approximately 4 h using chilled buffers. Gels were blotted onto PVDF Immobilon-P (Millipore) at 4 °C using a wet blotting system (BioRAD) with chilled buffers. For western blotting, primary antibodies were used at 1:1,000 and secondary antibodies were used at 1:5,000. Detection after primary antibody incubation used IRDye 680 goat anti-rabbit IgG or IRDye 800CW goat anti-mouse IgG, and was analysed via a Li-COR Odyssey Imager.

Immunofluorescence. For methanol fixation to preserve the overall cell shape, cells were spread onto microscope slides, and air-dried before fixation in methanol at −20 °C for at least 10 min. Cells were re-hydrated in PBS for 10 min before labelling, this being carried out as for paraformaldehyde fixed cells. For paraformaldehyde fixation, 2×10^6 cells were re-suspended in 100 μl vPBS (8 g l^{-1} NaCl, 0.22 g l^{-1} KCl, 2.27 g l^{-1} Na_2HPO_4, 0.41 g l^{-1} KH_2PO_4, 15.7 g l^{-1} sucrose, 1.8 g l^{-1} glucose; pH 7.4), and then an equal volume of 6% paraformaldehyde added. After 10 min the suspension was diluted to 5 ml in vPBS and settled onto poly-L-lysine-coated slides for 20 min and then washed with PBS. The cells were then permeabilized in 0.05% Triton X-100 for 10 min, blocked in 20% fetal calf serum (FCS) in vPBS for 45 min, and stained with the primary antibody (1:100) in 20% FCS in vPBS for >1 h.

Subsequent to washing in excess PBS, the cells were stained with the secondary antibody (1:500) in 20% FCS: vPBS for 1 h, washed in excess PBS 3×5 min after which the cellular DNA was stained with 1 μg ml^{-1} 4′,6-diamidino-2-phenylindole. Cells were mounted in MOWIOL containing phenylene diamine.

Flow cytometry. Between 2–5 $\times 10^6$ cells were fixed in 2% formaldehyde/0.05% glutaraldehyde for a minimum of 1 h at 4 °C. Subsequently, the cell suspension was pelleted and re-suspended in 200 μl of EP-procyclin antibody (Cedar Lane Laboratories) diluted 1:500 in 2% BSA in PBS. Cells were washed twice before being stained with the primary antibody, washed twice and stained with the secondary antibody (1:500). Flow cytometry data were analysed using FlowJO (Tree Star Inc.) software, with unstained cells, and cells stained with only the secondary antibody providing negative controls.

Image acquisition equipment and settings. Immunofluorescence microscopy images (Fig. 2c) were captured on a Zeiss axioskop 2 (Carl Zeiss microimaging) with a Prior Lumen 200 light source using a QImaging Retiga 2000R CCD camera; objectives were either Plan Neofluar ×63 (1.25 NA) or Plan Neofluar ×100 (1.30 NA). Images were captured via QImage (QImaging) and pseudocoloured using Adobe Photoshop CS. Confocal imaging (Figs 2a and 3c) used a Leica SP5 confocal laser scanning microscope, using ×63 oil immersion objective (NA = 1.4), with ×4.2 zoom. The green channel was imaged using a 488-nm argon laser, the red channel was imaged using a 543-nm helium/neon laser and the image was acquired at 1,024/1,024 voxels for *x*/*y* resolution with sequential optical sections of 0.54 μm in *z*-axis increments. The image was optimized by adjusting laser power and detector sensitivity to minimize bleaching and maintain a digital signal of between 0–255 to avoid signal loss or saturation. The final image was acquired using Volocity Software (Improvision Ltd.) Version 4.4.

Vol 459|14 May 2009|doi:10.1038/nature07983

ARTICLES

Select *Drosophila* glomeruli mediate innate olfactory attraction and aversion

Julia L. Semmelhack[1] & Jing W. Wang[1]

Fruitflies show robust attraction to food odours, which usually excite several glomeruli. To understand how the representation of such odours leads to behaviour, we used genetic tools to dissect the contribution of each activated glomerulus. Apple cider vinegar triggers robust innate attraction at a relatively low concentration, which activates six glomeruli. By silencing individual glomeruli, here we show that the absence of activity in two glomeruli, DM1 and VA2, markedly reduces attraction. Conversely, when each of these two glomeruli was selectively activated, flies showed as robust an attraction to vinegar as wild-type flies. Notably, a higher concentration of vinegar excites an additional glomerulus and is less attractive to flies. We show that activation of the extra glomerulus is necessary and sufficient to mediate the behavioural switch. Together, these results indicate that individual glomeruli, rather than the entire pattern of active glomeruli, mediate innate behavioural output.

The olfactory systems of phylogenetically diverse species have several common features[1,2], many of which are found in *Drosophila*. For example, each olfactory receptor neuron (ORN) expresses one or a few receptor genes that determine its odorant response profile[3–6], all ORNs expressing the same receptor genes project to the same glomerulus[5–8], and most output neurons send dendrites to a single glomerulus[9–11]. Thus, each glomerulus can be considered a functional unit. A single odorant typically activates several receptor types[12,13], and therefore elicits a distinct spatial pattern of activated glomeruli in the antennal lobe[14–16]. However, the mechanism by which these patterns are actually used to drive behavioural responses remains to be determined. It is possible that the whole pattern is necessary to elicit behavioural output. Alternatively, parts of the pattern, or even individual glomeruli, could be important for olfactory behaviours. This information from the antennal lobe can then be read out by higher brain centres, which probably integrate information from several sensory modalities to generate motor responses.

In contrast to the patterns of several glomeruli activated by most odorants, recent studies have identified two odorants that activate single glomeruli—CO_2 and the male-specific pheromone *cis*-vaccenyl acetate (cVA)—and trigger innate avoidance and female courtship receptivity, respectively[17–21]. By manipulating activity in the cognate receptor neurons, the activation of these single ORN channels was shown to be necessary and sufficient to produce the behaviour, suggesting that these receptors are hardwired to specific behavioural outputs[17,18,22]. These examples could be special cases because these odorants activate only one glomerulus, whereas most odorants excite several glomeruli. Furthermore, food odours contain many individual odorants[23], thus activating multiple glomeruli. Here we set out to study innate attraction to cider vinegar, a complex and highly attractive food odour, and to determine the role of individual glomeruli within the odour-evoked pattern.

Behavioural assay

Fruitflies are highly attracted to vinegar, which is associated with their favourite food source, rotting fruit[24]. To observe this innate attraction behaviour in individual flies, we used a four-field olfactometer design, which was recently applied to *Drosophila*[21]. By recording the outcome of several decisions in each fly, we were able to obtain a robust and reliable score even when using a relatively small number of flies. We measured attraction by observing single flies walking in a four-field arena, in which each quadrant received a separate air stream. When vinegar was added to one of the air streams, the fly spent most of its time in the corresponding quadrant (Fig. 1a). We recorded the location of the fly at 1-s intervals, and calculated a performance index by measuring the time spent in the odour quadrant. A fly that remained in the odour quadrant for the length of the assay scored 100%, whereas a fly that distributed its time equally among the four quadrants scored 0%, and a fly that spent no time in the odour quadrant scored −100%.

Using a concentration of 3 p.p.m. (isobutylene equivalents) vinegar, we saw an average performance index (PI) of 75% (Fig. 1b), which is consistent with previous results[21]. To verify that the behaviour is mediated by the olfactory system, we measured attraction in flies whose antennae had been amputated, and found that they were indifferent to vinegar ($PI = -6.7\%$, $n = 20$). Furthermore, we tested flies with a targeted mutation of *Or83b*. *Or83b* is expressed in 80% of all ORNs[8], and acts together with other olfactory receptors to generate responses to odorants[25–27]. We found that attraction was virtually abolished in *Or83b* mutant flies (Fig. 1b, c), with the distribution of control w^{1118} flies almost entirely separated from the *Or83b* mutant animals (Fig. 1d and Supplementary Fig. 1). In the absence of odours, control and mutant flies are distributed equally in all four quadrants (Supplementary Fig. 2), and *Or83b* mutant flies showed no impairment in CO_2 avoidance ($PI = -87 \pm 9\%$, mean ± s.e.m., $n = 12$), suggesting that their locomotion capability is normal. Thus, attraction in this assay requires ORNs, and the *Or83b* mutation provides a useful tool to link ORN activity with behavioural output.

Visualizing glomerular activity

We next determined which glomeruli are activated by vinegar. We used the genetically encoded calcium sensor G-CaMP to monitor activity in the antennal lobe using two-photon microscopy[15]. We imaged flies bearing the *GH146-Gal4* (also known as *P{GAL4}GH146*) and *UAS-GCaMP* (*P{UAS-G-CaMP}*) transgenes, which have G-CaMP expression in 83 out of 150 projection neurons[9,28,29]. Projection neurons are the output neurons of the antennal lobe; thus their responses to

[1]Neurobiology Section, Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA.

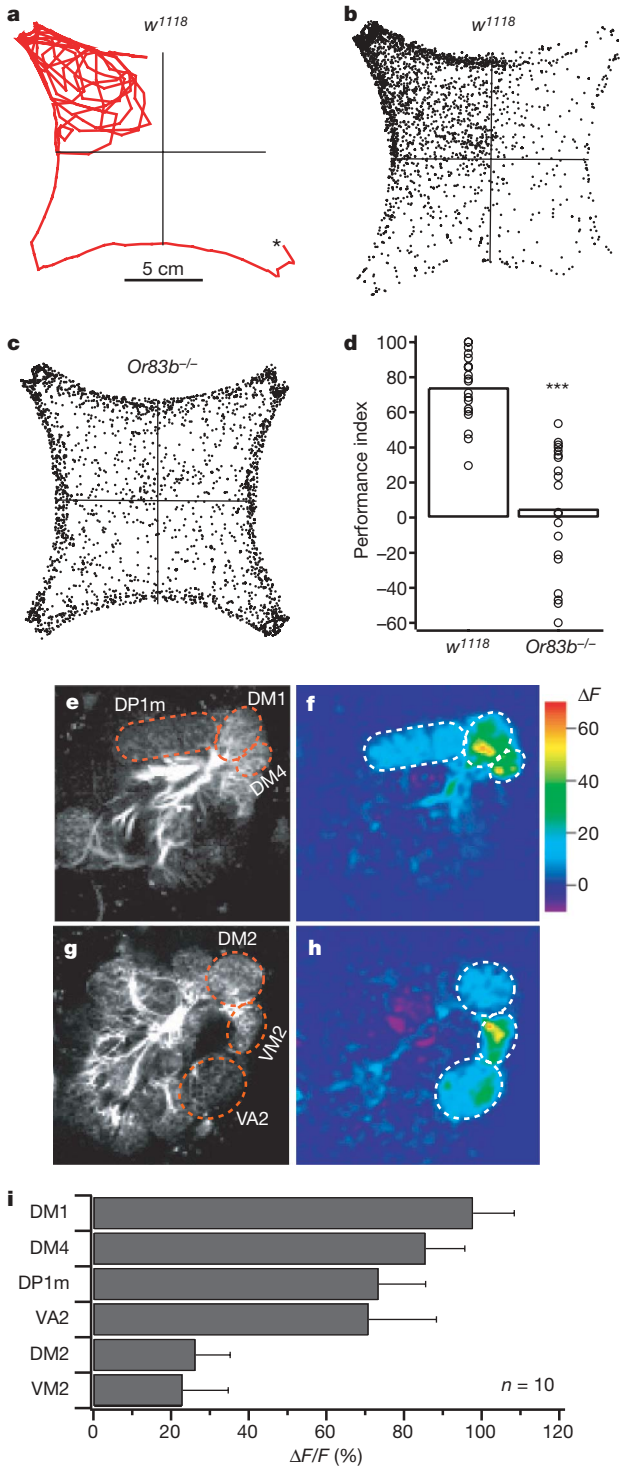


Figure 1 | Flies are robustly attracted to apple cider vinegar, which excites six glomeruli. **a**, Path of a single fly with 3 p.p.m. vinegar in the top left quadrant. **b**, **c**, Density plot of 20 w^{1118} (**b**) and $Or83b^{-/-}$ (**c**) flies. **d**, Performance index of w^{1118} and $Or83b^{-/-}$ flies. ***$P < 0.001$; *t*-test. **e**, **g**, Pre-stimulation images showing glomerular structure. The antennal lobe is roughly 65 μm in diameter. **f**, **h**, Responses to 3 p.p.m. vinegar in flies bearing the *GH146-Gal4* and *UAS-GCaMP* transgenes. **i**, Quantification of the change in fluorescence ($\Delta F/F$) for all six glomeruli over ten flies. Error bars indicate s.e.m.

odorants contain the information that is important for the behavioural response. We also imaged ORNs in flies bearing *Or83b-Gal4* and *UAS-GCaMP*, and found that the projection neuron response pattern is similar to the response of the ORNs (Supplementary Fig. 3), a result that is consistent with previous studies[14,15]. Although excitatory interglomerular connections do exist[30], recent studies have found that ORN input is the main determinant of projection neuron output[31,32].

From projection neuron and ORN imaging, we found that at 3 p.p.m. (the concentration used for the behavioural assay) vinegar elicited a response in 6 out of 34 glomeruli labelled by *GH146-Gal4*. In the most posterior plane of the antennal lobe, three glomeruli—DM1, DM4 and DP1M—responded quite robustly (Fig. 1e, f, i). On a more anterior plane, three more glomeruli—DM2, VA2 and VM2—also responded to varying degrees (Fig. 1g–i). Thus, at this behaviourally relevant concentration, vinegar excites six glomeruli. Although vinegar is a complex stimulus with many volatile components, previous studies have shown that several natural stimuli also elicit a surprisingly sparse response in the rodent olfactory bulb[33].

Two glomeruli relevant for attraction

To determine the role each activated glomerulus has in mediating the attraction to vinegar, we silenced each ORN channel in turn and addressed how that affected the attraction behaviour. Recently, a nearly complete map of ORN to glomerulus targeting was generated[5,6], so we were able to match five of the six activated glomeruli with their corresponding olfactory receptors (the receptor for DP1m remains unknown). *shibire*^{ts} is a temperature-sensitive mutant dynamin, which reversibly prevents neurotransmitter release at the non-permissive temperature (32 °C) by blocking endocytosis[34]. By generating flies bearing the *UAS-shi*^{ts} transgene and selective Or-Gal4 drivers, we should be able to silence five of the six glomeruli. Indeed, silencing individual ORN types resulted in a marked reduction in the activity of their cognate projection neurons, without affecting the non-cognate projection neuron response (Supplementary Fig. 4).

We found that when the Or42b neurons, which innervate the DM1 glomerulus, were silenced the attraction to vinegar was virtually eliminated (Fig. 2b, g). At the non-permissive temperature, the performance index for these flies was −4%, compared to 69% at the permissive temperature. To independently confirm this result, we have measured attraction behaviour in an *Or42b* mutant[35] and found a similar attraction deficit (PI = −18 ± 14%, $n = 18$). Silencing the Or92a neurons, which innervate the VA2 glomerulus, also had a marked effect on the behaviour, with the performance index declining to 50% at 32 °C (Fig. 2c, g). Flies with silenced DM4 and VM2 glomeruli showed normal attraction, as did all the genetic background controls (Fig. 2 and Supplementary Fig. 5). The deficits we observed when DM1 or VA2 were silenced suggest that these receptor neuron channels are required for the innate attraction behaviour, and could function as labelled lines for attraction. However, a model in which DM1 and VA2 are necessary for attraction in conjunction with other ORNs would also be consistent with these data.

We next asked whether individual receptor neuron channels could elicit attraction when activated alone. Because *Or83b* mutant flies lack a vital component of the olfactory signalling pathway and are non-responsive to vinegar, we reasoned that by restoring Or83b expression in specific ORNs, we could force vinegar to selectively activate a single Or83b-expressing glomerulus. Thus, we can determine what type of behavioural output each glomerulus would produce. We used *Or-Gal4* lines to drive expression of a *UAS-Or83b* transgene in *Or83b* mutant flies. Calcium imaging experiments confirmed that the rescue flies had normal olfactory responses in the corresponding ORNs (Supplementary Fig. 6). Notably, when the receptor neurons for either DM1 or VA2 were rescued, attraction was restored to normal levels (Fig. 3). These results indicate that it is activity in DM1 or VA2, and not the pattern of the six glomeruli, which is read out by higher brain centres to signal the attractiveness of the odour. The finding that VA2 activity is sufficient for attraction may seem inconsistent with the fact that DM1-silenced flies show no attraction to vinegar. However, VA2 may be more robustly activated in the rescue flies, because in the silencing experiments, activation of several remaining ORN channels could result in inhibition of VA2. Indeed, a recent study has shown that adding receptor channel inputs

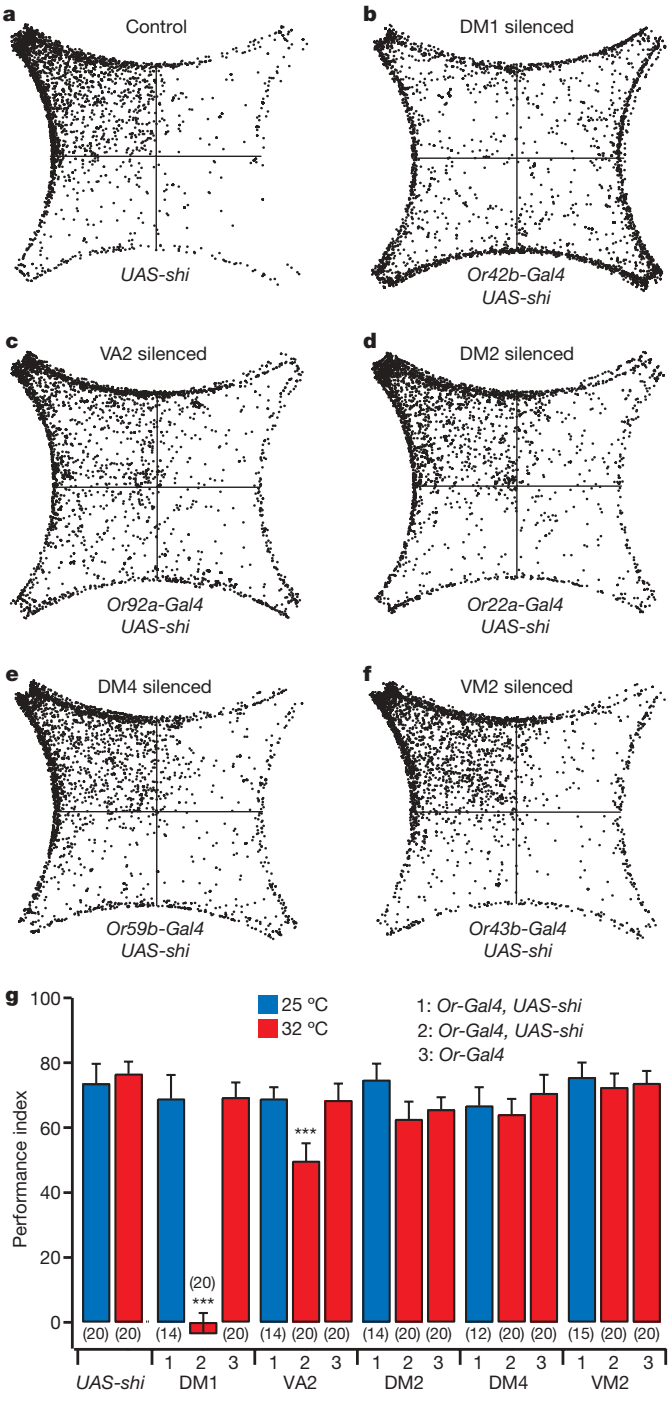


Figure 2 | Silencing DM1 or VA2 reduces attraction to 3 p.p.m. vinegar. **a–f**, Density plots composed of 20 flies each. **g**, Performance indices for flies bearing the *OrX-Gal4* and *UAS-shi^{ts}* transgenes at permissive and non-permissive temperatures. Analysis of variance (ANOVA) followed by Tukey's test was performed on PI values from flies of the experimental group at the permissive and non-permissive temperatures, and the corresponding genetic background controls at the non-permissive temperature. The number of flies is shown in parentheses. ***$P < 0.001$. Error bars indicate s.e.m.

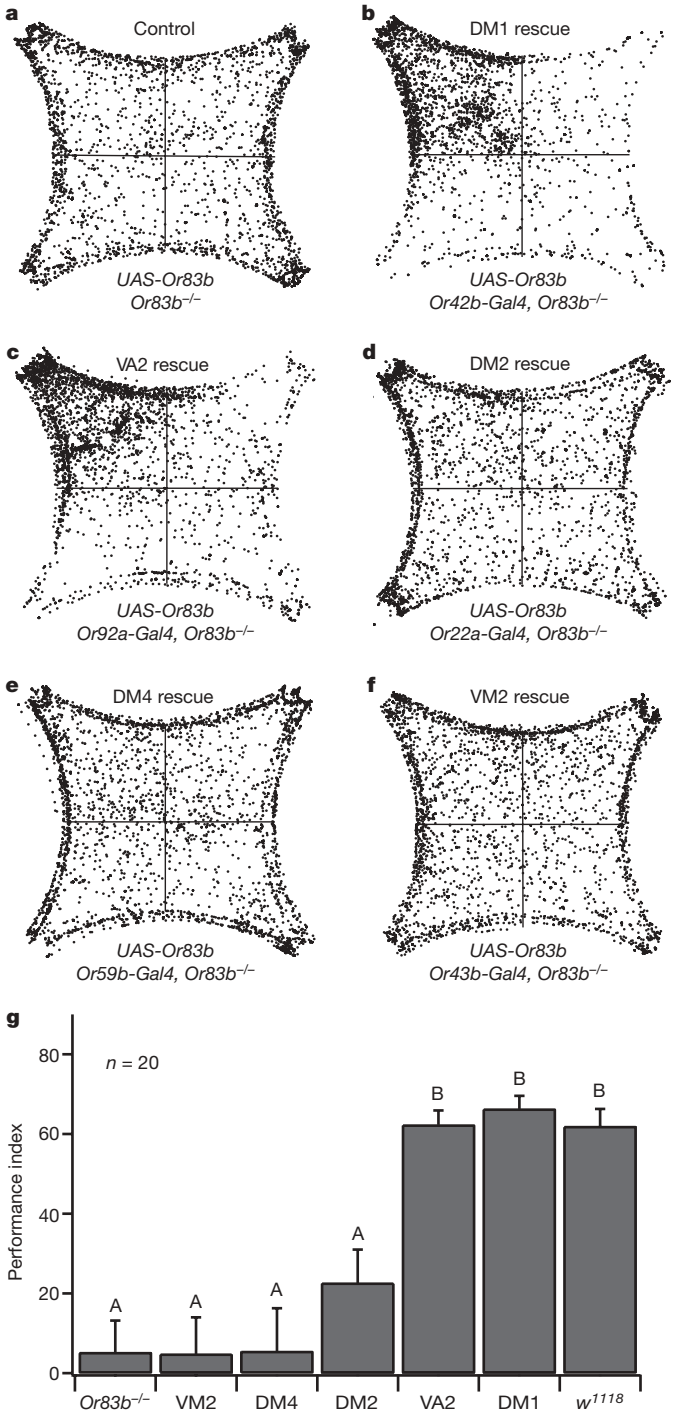


Figure 3 | Restoring *Or83b* in DM1 and VA2 ORNs returns attraction to control levels. **a–f**, Density plots of 20 flies responding to 3 p.p.m. vinegar. **g**, Performance indices of flies in which *Or83b* is selectively restored in individual ORN types. Comparisons between groups were made using ANOVA followed by Tukey's test. Significant differences ($P < 0.05$) are denoted by different letters. Error bars indicate s.e.m.

increases lateral inhibition, leading to a reduction in the projection neuron response[36].

Concentration-dependent behavioural switch

As odour concentration is increased, odours that are attractive at low concentrations often become less attractive or even repulsive[37]. Increasing the odorant concentration often recruits extra receptor neurons, and thus it has been proposed that the change in behaviour is mediated by the addition of these glomeruli to the ensemble of activated glomeruli[38], but this hypothesis has not been tested directly. It is also possible that the increased activation of the glomeruli that were active at the low concentration could mediate the change in behavioural output[39]. Alternatively, the new glomeruli could independently induce aversion.

We first measured the olfactory behaviour over a range of vinegar concentrations. As we increased the concentration, we observed a slight increase in attractiveness at 12 p.p.m. (Fig. 4a, b), but then a marked decrease in attractiveness at 32 p.p.m. with the performance index dropping to 9% (Fig. 4a, c). We wondered whether the change could

be due to the recruitment of additional glomeruli, and used calcium imaging to determine the difference in the pattern of glomeruli activated in response to 12 and 32 p.p.m. vinegar. We observed that the DM5 glomerulus, which showed no response at 12 p.p.m., was strongly activated at 32 p.p.m. (Fig. 4d, bottom row and 4e). All other glomeruli that were activated at 12 p.p.m. (DM1, DM4, DP1m, DM2, DM3, VM2 and VA2) showed small to moderate increases in response to 32 p.p.m. vinegar.

We next addressed whether DM5 could be responsible for the decrease in attraction to vinegar observed at 32 p.p.m. Therefore, we silenced the DM5 glomerulus by expressing *shibire*^{ts} in its cognate ORNs, which express Or85a. At the non-permissive temperature, we found that the performance index for 32 p.p.m. vinegar increased to 87% (Fig. 5a, c). In contrast, silencing DM1 resulted in repulsion towards 32 p.p.m. vinegar (Fig. 5c). Thus, the activation of DM5 is responsible for the decrease in attractiveness towards 32 p.p.m. vinegar.

In light of the above result, it is possible that the activation of DM5 alone mediates aversion, or that the activation of DM5 together with other specific glomeruli could mediate aversion. To distinguish between these models, we forced the stimulus to activate only DM5 by expressing Or83b in Or85a ORNs in the *Or83b* mutant background. We found that these flies were repulsed by 32 p.p.m. vinegar, whereas the *Or83b* mutant flies showed no preference or aversion to the odorant (Fig. 5b, c). In contrast, when DM1 was selectively activated by expression of *Or83b* in Or42b ORNs, flies were attracted to 32 p.p.m. vinegar (Fig. 5c and Supplementary Fig. 8). These findings suggest that the higher concentration of vinegar recruits an extra glomerulus that independently mediates aversion. When wild-type flies are exposed to 32 p.p.m. vinegar, the activation of an aversive glomerulus may counterbalance the activation of the two attractive glomeruli, resulting in a PI near zero.

If attraction and aversion are mediated by the activation of specific glomeruli, other odours that activate these glomeruli should give the same behavioural output. For example, an odour that excites DM1 should be attractive to flies in which DM1 ORNs are selectively activated, whereas an odour that selectively excites DM5 should be repulsive. We have identified an odorant, ethyl butyrate, that excites the DM1, DM2, VM2 and DM5 glomeruli (Supplementary Fig. 7), but has not been detected by gas chromatography in cider vinegar[40]. When we selectively restored function in DM1 ORNs, we found that ethyl butyrate triggered attraction behaviour, with a PI of 65%. Conversely, when we selectively restored function in the DM5 ORNs, the result was an aversion to ethyl butyrate, with a PI of −34% (Fig. 6a). These results indicate that the activation of DM1 or DM5 by any odour should be sufficient for attraction and aversion, respectively.

If specific glomeruli are hardwired to generate attraction and aversion behaviour, activation of ectopically expressed receptors should give a similar behavioural output. We predict that expression of the *Or22a* receptor in Or85a ORNs, which project to DM5, should make these neurons sensitive to lower concentrations of vinegar, and bias the behaviour towards aversion. Indeed, these flies show a marked reduction in the PI value in response to 12 p.p.m. vinegar (Fig. 6b), indicating that it is activity in the DM5 ORNs, rather than activation of a particular receptor, that biases the behaviour towards aversion.

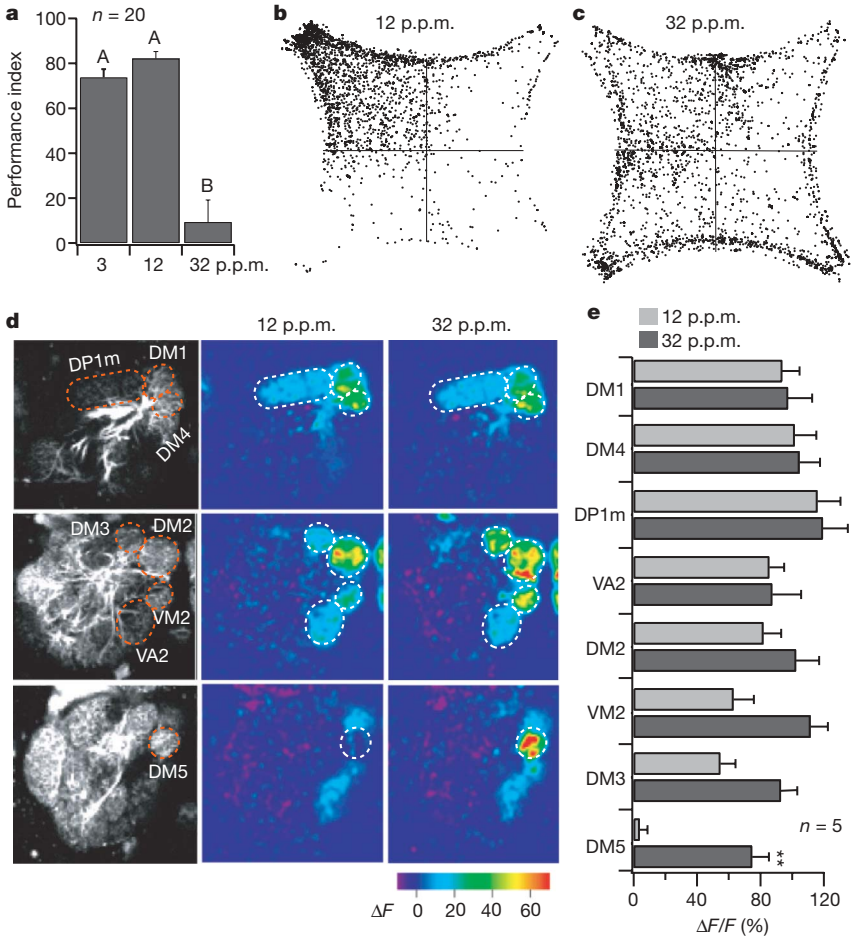


Figure 4 | Vinegar becomes less attractive and activates an additional glomerulus at high concentrations. **a**, Performance indices of w^{1118} flies at various concentrations of vinegar. PI values were compared using ANOVA followed by Tukey's test. Significant differences ($P < 0.05$) are denoted by different letters. **b**, **c**, Density plots of w^{1118} behaviour in response to 12 p.p.m. (**b**) and 32 p.p.m. (**c**) vinegar. **d**, Responses to 12 p.p.m. and 32 p.p.m. vinegar in flies bearing the *GH146-Gal4* and *UAS-GCaMP* transgenes. The antennal lobe is roughly 65 μm in diameter. **e**, The average change in fluorescence ($\Delta F/F$) is shown. $^{**}P < 0.01$; *t*-test. Error bars indicate s.e.m.

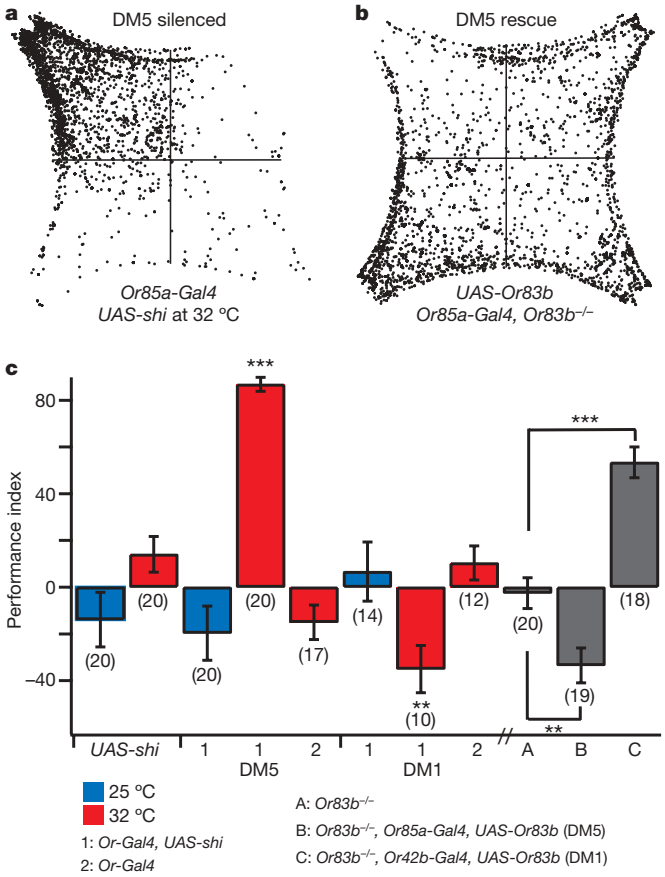


Figure 5 | DM5 mediates the decrease in attraction in response to 32 p.p.m. vinegar. a, Density plot of 20 flies in which the DM5 ORNs are silenced. **b**, Density plot of 20 DM5 rescue flies. **c**, Behavioural responses to 32 p.p.m. vinegar for flies in which DM5 and DM1 are silenced and selectively rescued. For silencing experiments, we performed the same statistical analysis as in Fig. 2. DM5 rescue and DM1 rescue flies were compared to $Or83b^{-/-}$ flies by *t*-test. $^{**}P < 0.01$, $^{***}P < 0.001$.

Discussion

Previous studies have shown that in certain cases, olfactory behaviours are elicited by dedicated receptor channels or labelled lines[17,18]. In this study, we demonstrate that innate attraction to a complex food odour is similarly mediated by a few of the activated glomeruli. However, it is possible that other glomeruli not activated by cider vinegar could also mediate innate attraction to other food odours. A recent study of olfactory behaviour in *Drosophila* larvae also addressed how receptor activation leads to behavioural output, and found that the responses of five ORNs to a panel of odorants can be used to generate a model that accounts for 81% of the variation in olfactory behaviour[39]. Selective activation of these ORNs should generate robust innate attraction or aversion. In fact, the Or42a ORN, one of the five critical ORNs, has been shown to be sufficient for attraction behaviour[41].

Furthermore, we show that the decrease in attractiveness in response to a higher concentration of vinegar is due to the activation of an additional glomerulus. It is a common feature of olfactory perception that most odours become less pleasant and eventually repellent as their intensity is increased[37], a phenomenon that has also been observed in *Drosophila*[23,42,43]. The recruitment of further glomeruli has been proposed as a mechanism to mediate this change in behavioural output[38]. A recent paper has suggested that different levels of activation in the same ORNs could generate qualitatively different behavioural responses[39]. Here we found that a glomerulus recruited by a high concentration of vinegar, DM5, has an important role in the behavioural switch. Silencing and selective activation experiments show that DM5 is necessary and sufficient for the behavioural switch.

The present results indicate that certain olfactory receptor neurons in *Drosophila* are genetically hardwired to generate robust innate olfactory attraction or avoidance behaviour, an organizing principle that has been observed in several chemosensory systems[17,18,44,45]. In the fly, projection neurons receive input from ORNs and send axons to the mushroom body and lateral horn. Further studies should shed light on the mechanism by which these centres generate the behaviours we observe.

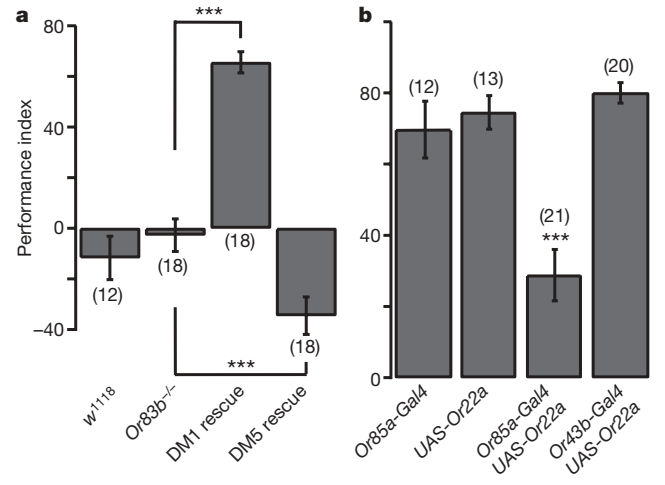


Figure 6 | DM1 and DM5 mediate attraction and aversion in response to ethyl butyrate. a, Performance indices in response to 7 p.p.m. ethyl butyrate for DM1 and DM5 rescue flies. $^{***}P < 0.001$; *t*-test. **b**, Ectopic expression of Or22a in Or85a ORNs reduced attraction to 12 p.p.m. vinegar. PI values were compared using ANOVA followed by Tukey's test. $^{***}P < 0.001$.

METHODS SUMMARY

Behavioural assay. An existing behavioural model was modified to measure the response of single flies to odours[21]. The four-field olfactometer consisted of a four-pointed star-shaped arena. Air flow was maintained by vacuum suction, such that air entered each quadrant at a rate of 200 ml min^{-1}, after passing through a 100-ml bottle. Female flies that had been starved for 50 h were used. After the addition of an odorant to one quadrant, the fly's location was measured once per second. The performance index is defined as $(2p^{1/2} - 1) \times 100\%$, in which p is the fraction of time the fly spends in the odour quadrant between 50 and 250 s after odour application.

Odour stimuli. Odour concentration was measured using a photoionization detector (Rae Systems, MiniRAE 2000) and an air flow of 200 ml min^{-1} through a 100-ml bottle containing the odorant. As the conversion factor to determine the exact concentration of cider vinegar volatiles is unknown, we express the concentration in isobutylene equivalents. The 3 p.p.m. concentration of vinegar corresponds to 40 µl of a 1:2 dilution of apple cider vinegar with water on filter paper. Twelve parts per million is 80 µl vinegar, 32 p.p.m. is 1 ml vinegar, and 7 p.p.m. ethyl butyrate came from 40 µl of a 1:1,000 dilution of ethyl butyrate in mineral oil. The odour source was replenished for each experiment. Odour concentrations stayed constant over the time course of an experiment.

G-CaMP imaging experiments. Calcium imaging was performed as described[15,46] except that the air-flow rate was 200 ml min^{-1}. Odorants were administered from 100-ml bottles as described earlier, and stimuli were given for 2 s.

Transgenic flies. The following fly stocks were used: *Or42b-Gal4*, *Or43b-Gal4*, *Or92a-Gal4*, *Or22a-Gal4* and *Or92a-Gal4* (ref. 5), *Or59b-Gal4* (ref. 6), *UAS-Or22a*, *UAS-Or83b*, *Or83b*-targeted deletion ($Or83b^2$)[26], $UAS\text{-}shibire^{ts}$ (ref. 34), *UAS-GCaMP* (ref. 15), *GH146-Gal4* (ref. 29), *GH146-LexAGAD*[47] and *LexAop-GCaMP-IRES-GCaMP*[46].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 September 2008; accepted 12 March 2009.
Published online 26 April 2009.

1. Hildebrand, J. G. & Shepherd, G. M. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
2. Bargmann, C. I. Comparative chemosensation from receptors to ecology. *Nature* **444**, 295–301 (2006).

3. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96,** 725–736 (1999).
4. Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G. & Carlson, J. R. Coexpression of two functional odor receptors in one neuron. *Neuron* **45,** 661–666 (2005).
5. Fishilevich, E. & Vosshall, L. B. Genetic and functional subdivision of the *Drosophila* antennal lobe. *Curr. Biol.* **15,** 1548–1553 (2005).
6. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15,** 1535–1547 (2005).
7. Gao, Q., Yuan, B. & Chess, A. Convergent projections of *Drosophila* olfactory neurons to specific glomeruli in the antennal lobe. *Nature Neurosci.* **3,** 780–785 (2000).
8. Vosshall, L. B., Wong, A. M. & Axel, R. An olfactory sensory map in the fly brain. *Cell* **102,** 147–159 (2000).
9. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109,** 229–241 (2002).
10. Marin, E. C., Jefferis, G. S., Komiyama, T., Zhu, H. & Luo, L. Representation of the glomerular olfactory map in the *Drosophila* brain. *Cell* **109,** 243–255 (2002).
11. Stocker, R. F., Lienhard, M. C., Borst, A. & Fischbach, K. F. Neuronal architecture of the antennal lobe in *Drosophila melanogaster. Cell Tissue Res.* **262,** 9–34 (1990).
12. de Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30,** 537–552 (2001).
13. Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117,** 965–979 (2004).
14. Ng, M. *et al.* Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron* **36,** 463–474 (2002).
15. Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112,** 271–282 (2003).
16. Fiala, A. *et al.* Genetically expressed cameleon in *Drosophila melanogaster* is used to visualize olfactory information in projection neurons. *Curr. Biol.* **12,** 1877–1884 (2002).
17. Suh, G. S. *et al.* A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila. Nature* **431,** 854–859 (2004).
18. Kurtovic, A., Widmer, A. & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446,** 542–546 (2007).
19. Clyne, P., Grant, A., O'Connell, R. & Carlson, J. R. Odorant response of individual sensilla on the *Drosophila* antenna. *Invert. Neurosci.* **3,** 127–135 (1997).
20. Ha, T. S. & Smith, D. P. A pheromone receptor mediates 11-cis-vaccenyl acetate-induced responses in *Drosophila. J. Neurosci.* **26,** 8727–8733 (2006).
21. Faucher, C., Forstreuter, M., Hilker, M. & de Bruyne, M. Behavioral responses of *Drosophila* to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context. *J. Exp. Biol.* **209,** 2739–2748 (2006).
22. Suh, G. S. *et al.* Light activation of an innate olfactory avoidance response in *Drosophila. Curr. Biol.* **17,** 905–908 (2007).
23. Stensmyr, M. C., Giordano, E., Balloi, A., Angioy, A. M. & Hansson, B. S. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J. Exp. Biol.* **206,** 715–724 (2003).
24. Zhu, J., Park, K. C. & Baker, T. C. Identification of odors from overripe mango that attract vinegar flies, *Drosophila melanogaster. J. Chem. Ecol.* **29,** 899–909 (2003).
25. Sato, K. *et al.* Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452,** 1002–1006 (2008).
26. Larsson, M. C. *et al. Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43,** 703–714 (2004).
27. Wicher, D. *et al. Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452,** 1007–1011 (2008).
28. Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila. Annu. Rev. Neurosci.* **30,** 505–533 (2007).
29. Stocker, R. F., Heimbeck, G., Gendre, N. & de Belle, J. S. Neuroblast ablation in *Drosophila* P[GAL4] lines reveals origins of olfactory interneurons. *J. Neurobiol.* **32,** 443–456 (1997).
30. Shang, Y., Claridge-Chang, A., Sjulson, L., Pypaert, M. & Miesenbock, G. Excitatory local circuits and their implications for olfactory processing in the fly antennal lobe. *Cell* **128,** 601–612 (2007).
31. Olsen, S. R., Bhandawat, V. & Wilson, R. I. Excitatory interactions between olfactory processing channels in the *Drosophila* antennal lobe. *Neuron* **54,** 89–103 (2007).
32. Root, C. M., Semmelhack, J. L., Wong, A. M., Flores, J. & Wang, J. W. Propagation of olfactory information in *Drosophila. Proc. Natl Acad. Sci. USA* **104,** 11826–11831 (2007).
33. Lin, D. Y., Shea, S. D. & Katz, L. C. Representation of natural stimuli in the rodent main olfactory bulb. *Neuron* **50,** 937–949 (2006).
34. Kitamoto, T. Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive *shibire* allele in defined neurons. *J. Neurobiol.* **47,** 81–92 (2001).
35. Bellen, H. J. *et al.* The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167,** 761–781 (2004).
36. Asahina, K., Louis, M., Piccinotti, S. & Vosshall, L. B. A circuit supporting concentration-invariant odor perception in *Drosophila. J. Biol.* **8,** 9 (2009).
37. Laing, D. G., Panhuber, H. & Baxter, R. I. Olfactory properties of amines and *n*-butanol. *Chem. Senses* **3,** 149–166 (1978).
38. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96,** 713–723 (1999).
39. Kreher, S. A., Mathew, D., Kim, J. & Carlson, J. R. Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59,** 110–124 (2008).
40. Aurand, L. W., Singleton, J. A., Bell, T. A. & Etchells, J. L. Volatile components in the vapors of natural and distilled vinegars. *J. Food Sci.* **31,** 172–177 (1966).
41. Fishilevich, E. *et al.* Chemotaxis behavior mediated by single larval olfactory neurons in *Drosophila. Curr. Biol.* **15,** 2086–2096 (2005).
42. Heimbeck, G., Bugnon, V., Gendre, N., Keller, A. & Stocker, R. F. A central neural circuit for experience-independent olfactory and courtship behavior in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA* **98,** 15336–15341 (2001).
43. Wang, Y. *et al.* Blockade of neurotransmission in *Drosophila* mushroom bodies impairs odor attraction, but not repulsion. *Curr. Biol.* **13,** 1900–1904 (2003).
44. Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans. Cell* **91,** 161–169 (1997).
45. Zhao, G. Q. *et al.* The receptors for mammalian sweet and umami taste. *Cell* **115,** 255–266 (2003).
46. Root, C. M. *et al.* A presynaptic gain control mechanism fine-tunes olfactory behavior. *Neuron* **59,** 311–321 (2008).
47. Lai, S. L., Awasaki, T., Ito, K. & Lee, T. Clonal analysis of *Drosophila* antennal lobe neurons: diverse neuronal architectures in the lateral neuroblast lineage. *Development* **135,** 2883–2893 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank D. Anderson, R. Axel, C. Zuker and M. Gallio for comments on the manuscript. We thank M. Hilker for help with the design of the four-field olfactometer. We thank C. Root for generating the data presented in Supplementary Fig. 4. We thank L. Vosshall, B. Dickson and T. Lee for providing fly stocks. This work was partially supported by a research grant from the Whitehall Foundation to J.W.W. and a grant from the National Institute of Deafness and other Communication Disorders to J.W.W. (R01DC009597). J.W.W. is a Beckman Young Investigator, a Hellman Faculty scholar, and a Searle scholar.

Author Contributions J.L.S. and J.W.W. designed experiments, analysed the data, and wrote the manuscript. J.L.S. carried out the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.W.W. (jw800@ucsd.edu).

METHODS

Behavioural assay. An existing behavioural model was modified to measure the response of single flies to odours[21]. As previously described, the four-field olfactometer consisted of a four-pointed star-shaped arena 30 cm across diagonally and 1 cm deep, covered by a glass plate. Air flow was maintained by vacuum suction such that air entered each quadrant at a rate of 200 ml min^{-1}, after passing through a 100-ml bottle. Only female flies were used, and at the time of the assay the flies were 4-days-old and had been starved for 50 h in a vial with a wet kimwipe. After a single fly was introduced into the chamber, its speed was measured for 100 s, and only flies with an average speed between 0.5 and 1.0 cm s^{-1} were used. At the start of the assay, one of the empty 100-ml bottles was replaced with an odour-containing bottle. The fly's location was measured once per second using a Logitech quickcam and Labview software (National Instruments). The chamber was illuminated by a panel of light-emitting diodes (660 nm). Light reflected from the glass plate was eliminated by polarizing optics. The performance index is defined as $(2p^{1/2} - 1) \times 100\%$, in which p is the fraction of time the fly spends in the odour quadrant during the period between 50 and 250 s after odour application. Thus, if the fly is in the odour quadrant for the entire time window, $P = 1$ and the performance index is 100%, whereas if the fly avoids the odour quadrant entirely, $P = 0$ and the performance index will be -100%. Except for the *shibire*^{ts} non-permissive temperature experiments (which were performed at 32 °C), all behavioural experiments were performed at 25 °C and 70% humidity. Data were analysed using Igor Pro (Wavemetrics) and a custom macro. The Jarqe-Bera test was used to verify that the data were normally distributed. Density plots show data collected between 50 and 250 s for 20 flies. Each dot indicates one fly spending one second at that location. Odour application was alternated among the four quadrants, and the density plots were created by rotating the positional data so that the odour quadrant becomes the top left quadrant.

Vol 459|14 May 2009|doi:10.1038/nature08004

LETTERS

Episodic formation of cometary material in the outburst of a young Sun-like star

P. Ábrahám[1], A. Juhász[2], C. P. Dullemond[2], Á. Kóspál[3], R. van Boekel[2], J. Bouwman[2], Th. Henning[2], A. Moór[1], L. Mosoni[1,2], A. Sicilia-Aguilar[2] & N. Sipos[1]

The Solar System originated in a cloud of interstellar gas and dust. The dust is in the form of amorphous silicate particles[1,2] and carbonaceous dust. The composition of cometary material, however, shows that a significant fraction of the amorphous silicate dust was transformed into crystalline form during the early evolution of the protosolar nebula[3]. How and when this transformation happened has been a question of debate, with the main options being heating by the young Sun[4,5] and shock heating[6]. Here we report mid-infrared features in the outburst spectrum of the young Sun-like star EX Lupi that were not present in quiescence. We attribute them to crystalline forsterite. We conclude that the crystals were produced through thermal annealing in the surface layer of the inner disk by heat from the outburst, a process that has hitherto not been considered. The observed lack of cold crystals excludes shock heating at larger radii.

The year 2008 brought a rare opportunity to study high-temperature dust processing on a human timescale in a cosmic laboratory. The 'experiment' took place in the circumstellar disk of EX Lupi, an M0 star that is the prototype of a class of young eruptive stars named EXors[7]. These objects are defined by their large, repetitive outbursts, which are attributed to temporarily increased mass accretion from the circumstellar disk onto the star[8]. Such outbursts represent the most intense accretion episodes in assembling the final stellar mass. In January 2008, EX Lupi entered one of its largest outbursts, brightening by a factor of ~100 and reaching a maximum brightness of 8 mag in visual light[9,10]. We observed EX Lupi in the 5.2–37-μm wavelength range with the Infrared Spectrograph on board NASA's Spitzer Space Telescope, on 2008 April 21. EX Lupi was already slowly fading after its peak brightness in 2008 February, but was still a factor of 30 brighter in visual light than in quiescence.

Comparing our EX Lupi spectrum with a pre-outburst measurement from the Spitzer archive obtained in 2005, we observed a significant change. In the 8–12-μm spectral range (Fig. 1), the silicate profile in quiescence exhibited a triangular shape, similar to that of the amorphous interstellar grains[1,11] (Fig. 1a, b). By contrast, in outburst (Fig. 1c) we clearly detected several narrower spectral features, which we identified as crystalline silicates, on top of the broad peak of amorphous silicates. The sharp peak at 10 μm and the shoulder at 11.3 μm suggest that forsterite, the magnesium-rich form of olivine, dominates the observed crystal population[12,13]. The appearance of a weaker peak at 16 μm (Supplementary Information) supports this conclusion. At longer wavelengths, no other crystalline features are present in the spectrum. The observed crystalline features are similar to those present in comet spectra[3,14] (Fig. 1d) and in a number of protoplanetary disks[15,16]. The remaining differences between EX Lupi and the cometary spectra can be related to different temperatures and different relative abundances of dust components.

Because the quiescent spectrum has no crystalline silicate features, the appearance of crystalline features in the EX Lupi outburst strongly suggests that we witnessed on-going crystal formation.

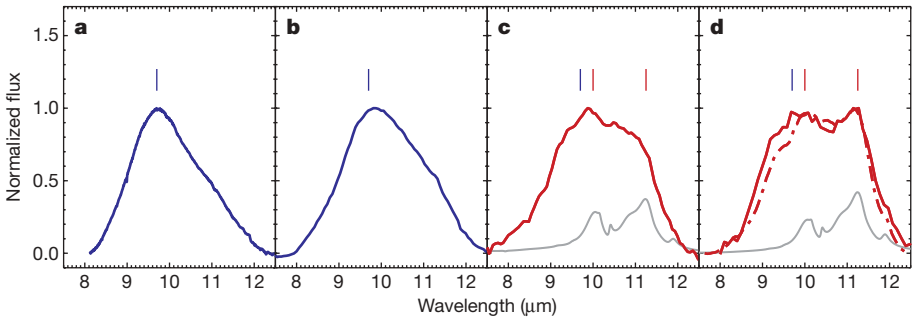


Figure 1 | Silicate emission in the 8–12-μm range. **a**, Spectrum of interstellar grains measured in the direction of the Galactic Centre[1]. **b**, Spitzer Infrared Spectrograph spectrum of EX Lupi, obtained on 2005 March 18, in quiescent phase. **c**, Our Spitzer spectrum of EX Lupi, obtained on 2008 April 21, in the middle of the present outburst. **d**, Red line, ground-based spectrum of Comet 1P/Halley[3]; dash–dot line, Spitzer spectrum of the ejecta from Comet 9P/Tempel 1 during the Deep Impact experiment[14] (available in the Spitzer archive). After a linear continuum removal, the spectra were normalized to their peak values. In **a**, we see the characteristic triangular shape profile attributed to amorphous silicate grains[1]; the vertical blue dash at 9.7 μm (repeated in all panels) corresponds to the peak wavelength of the amorphous silicate profile as measured in the laboratory[11]. In **b**, the EX Lupi spectrum closely resembles the amorphous profile, with some slight excess on the long-wavelength side. In **c**, peaks and shoulders due to crystalline silicates can be identified. Peak wavelengths of forsterite at 10.0 and 11.2 μm, as measured in laboratory experiments[12,13], are marked by red dashes. The grey curves in **c** and **d** display the emissivity curve of pure forsterite[13], assuming representative silicate grain temperatures of 1,250 K and 300 K, respectively. Panel **d** shows that the same crystalline features can be observed in cometary spectra.

[1]Konkoly Observatory of the Hungarian Academy of Sciences, PO Box 67, 1525 Budapest, Hungary. [2]Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany. [3]Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands.

Alternative explanations, such as the illumination of existing crystals residing in outer disk areas or the stirring up of crystals from the disk mid-plane can be excluded by modelling (Supplementary Information). The relative strengths of the 10- and 11.3-μm crystalline features, and the lack of spectral features beyond 16 μm, imply that the new crystals are hot and were formed in a high-temperature process.

We estimated a temperature range for this process using our recent radiative-transfer modelling of EX Lupi in quiescence[17], which assumes a circumstellar disk geometry encircling an inner dust-free hole of radius 0.2 AU (Supplementary Information, section 1.3). Owing to this hole, the temperature in the quiescent disk was almost everywhere below 700 K. The quiescent spectrum indicates that no noticeable crystal formation has occurred at these temperatures. The disk temperature in outburst was simulated by increasing the luminosity of the central source by a factor of ten, estimated from the flux increase at the highest-frequency part of the Spitzer spectra at around 5 μm. An assumed black-body spectrum of 6,800 K, typical of outbursting stars[8], accounted for the higher temperature of the source during eruption. The modelling revealed that in outburst a significant disk area became hotter than 700 K, but that its temperature was almost everywhere below 1,500 K, the vaporization threshold of silicate particles. Our observations point to a crystallization mechanism that works efficiently between 700 and 1,500 K in the protoplanetary environment.

Laboratory experiments suggest that this mechanism is thermal annealing[18,19]. According to laboratory measurements, above 1,000 K annealing occurs on very short timescales, of seconds to hours[18], fitting well within the observed timescale of the EX Lupi outburst. The radius of this crystal formation zone, 0.5 AU, is comparable to that of the terrestrial-planet region in the Solar System. The high temperature of the crystals excludes their formation in shock fronts at radii of several astronomical units[6], because grains behind the shock would quickly cool down and produce observable spectral features at wavelengths longer than 20 μm.

Our observations revealed another interesting process acting in the surface layer of the disk. In 1955–1956, EX Lupi underwent a major eruption very similar to the one discussed here[7], and it is likely that a similar amount of crystalline silicate formed then (Supplementary Information, section 1.4). However, by 2005, the date of the quiescent Spitzer spectrum, the crystalline spectral features had vanished, indicating that the crystals disappeared from the disk surface layer in less than 50 years. Possible explanations for this fast removal process, discussed further in the Supplementary Information, are vertical mixing that transports the crystals into the disk interior[4,20], inward surface flow that may accrete them onto the star[20] and amorphization by cosmic rays[21] or X-rays[22]. We note that in the similar, but even more violent, eruptions of FU Orionis-type young stars, crystals were searched for but not detected, and their lack speculated to be the effect of vertical mixing[23].

Our detections demonstrate that crystalline silicate grains can both form in and disappear from the surface layer of the disk within months to decades. Thus, we predict that multi-epoch measurements of the crystallinity level in EX Lupi would provide fluctuating results, depending on the actual activity level of the star and, in particular, on the time elapsed since its last major outburst. Therefore, the observed crystallinity is not a useful indicator of the dust composition in the disk interior, which would evolve much more gradually. Also, the observed dominance of forsterite may not be representative of the disk interior. Similar conclusions may hold for many other young stars. Pre-main-sequence evolution is generally accompanied by optical–infrared variability[24,25], and a large fraction—if not the majority—of young stars frequently change their luminosity by factors of <10. Thus, the observed crystallinity may vary considerably, and randomly, among stars of similar mass and age, and correlation of its value with the stellar parameters may be weaker and less informative than previously expected.

The observations of the outburst of EX Lupi point to a new mechanism of crystal formation in protoplanetary disks: episodic surface crystallization. It is now an observationally established crystal-forming mechanism acting in the inner disk region, supplementing crystal formation due to accretion heat in the disk mid-plane[4] or shocks[6]. We suggest that such crystallization events occur in the life of most young stars. With an average interval of 50 years, EX Lupi may still undergo several thousand such eruptions before the end of its early evolution. The current picture confines crystallization to the very early phases of pre-main-sequence evolution[4]. Our findings show that crystallization episodes can also continue in later phases, when accretion has significantly dropped (apart from the episodic outbursts), covering a significant part of the pre-main-sequence evolution. Moreover, this explanation may work even in disks with large inner holes (such as EX Lupi), where the mid-plane temperature will never be high enough for crystallization.

Our proposal for episodic crystallization is in line with recent observations which imply that there must be another grain crystallization mechanism uncorrelated with the steady mass-accretion rate, stellar luminosity, disk mass or disk/star mass ratio[26]. Although in a single outburst only the thin surface layer of the disk is crystallized, in EX Lupi we saw that subsequent major outbursts always transform a new layer of amorphous grains into crystals, potentially enriching the disk interior through vertical mixing. In any case, a fraction of the crystals may be mixed outwards, and may contribute to the build-up of protocomets. Assuming that a similar process occurred in the proto-Solar System, crystalline grains in comets and meteorites might be messengers of past eruptions, having been formed in a crucible around the outbursting young Sun.

Received 24 November 2008; accepted 19 March 2009.

1. Kemper, F., Vriend, W. J. & Tielens, A. G. G. M. The absence of crystalline silicates in the diffuse interstellar medium. *Astrophys. J.* **609,** 826–837 (2004).
2. Li, M. P., Zhao, G. & Li, A. On the crystallinity of silicate dust in the interstellar medium. *Mon. Not. R. Astron. Soc.* **382,** L26–L29 (2007).
3. Hanner, M. S., Lynch, D. K. & Russell, R. W. The 8–13 micron spectra of comets and the composition of silicate grains. *Astrophys. J.* **425,** 274–285 (1994).
4. Gail, H.-P. Radial mixing in protoplanetary accretion disks. I. Stationary disc models with annealing and carbon combustion. *Astron. Astrophys.* **378,** 192–213 (2001).
5. Bockelée-Morvan, D., Gautier, D., Hersant, F., Huré, J.-M. & Robert, F. Turbulent radial mixing in the solar nebula as the source of crystalline silicates in comets. *Astron. Astrophys.* **384,** 1107–1118 (2002).
6. Harker, D. E. & Desch, S. J. Annealing of silicate dust by nebular shocks at 10 AU. *Astrophys. J.* **565,** L109–L112 (2002).
7. Herbig, G. H. Eruptive phenomena in early stellar evolution. *Astrophys. J.* **217,** 693–715 (1977).
8. Hartmann, L. & Kenyon, S. J. The FU Orionis phenomenon. *Annu. Rev. Astron. Astrophys.* **34,** 207–240 (1996).
9. Jones, A. F. A. L. EX Lupi. *Central Bureau Electronic Telegrams* **1217** (2008).
10. Kóspál, Á. *et al.* The extreme outburst of EX Lupi in 2008: optical spectra and light curve. *Inf. Bull. Var. Stars* **5819,** 1–4 (2008).
11. Dorschner, J., Begemann, B., Henning, Th, Jaeger, C. & Mutschke, H. Steps toward interstellar silicate mineralogy. II. Study of Mg-Fe-silicate glasses of variable composition. *Astron. Astrophys.* **300,** 503–520 (1995).
12. Jaeger, C. *et al.* Steps toward interstellar silicate mineralogy. IV. The crystalline revolution. *Astron. Astrophys.* **339,** 904–916 (1998).
13. Koike, C. *et al.* Compositional dependence of infrared absorption spectra of crystalline silicate. II. Natural and synthetic olivines. *Astron. Astrophys.* **399,** 1101–1107 (2003).
14. Lisse, C. M. *et al.* Spitzer spectral observations of the Deep Impact ejecta. *Science* **313,** 635–640 (2006).
15. Bouwman, J. *et al.* Processing of silicate dust grains in Herbig Ae/Be systems. *Astron. Astrophys.* **375,** 950–962 (2001).
16. van Boekel, R. *et al.* The building blocks of planets within the 'terrestrial' region of protoplanetary disks. *Nature* **432,** 479–482 (2004).
17. Sipos, N. *et al.* The optical–infrared properties of EX Lupi in quiescent phase. *Astron. Astrophys.* (submitted).
18. Hallenbeck, S. L., Nuth, J. A. & Daukantas, P. L. Mid-infrared spectral evolution of amorphous magnesium silicate smokes annealed in vacuum: comparison to cometary spectra. *Icarus* **131,** 198–209 (1998).
19. Colangeli, L. *et al.* The role of laboratory experiments in the characterisation of silicon-based cosmic material. *Astron. Astrophys. Rev.* **11,** 97–152 (2003).
20. Ciesla, F. J. Two-dimensional transport of solids in viscous protoplanetary disks. *Icarus* **200,** 655–671 (2009).

21. Bringa, E. M. *et al.* Energetic processing of interstellar silicate grains by cosmic rays. *Astrophys. J.* **662,** 372–378 (2007).
22. Glauser, A. M. *et al.* Formation of crystalline dust grains in protoplanetary disks: observational evidence for the destructive effect of X-rays, in *Proc. 5th Spitzer Conf. "New Light on Young Stars: Spitzer's View of Circumstellar Disks"* ⟨www.ipac.caltech.edu/spitzer2008/posters/AdrianGlauser_Spitzer2008Conference.pdf⟩ (2008).
23. Quanz, S. P. *et al.* Evolution of dust and ice features around FU Orionis objects. *Astrophys. J.* **668,** 359–383 (2007).
24. Herbst, W., Herbst, D. K., Grossman, E. J. & Weinstein, D. Catalogue of UBVRI photometry of T Tauri stars and analysis of the causes of their variability. *Astron. J.* **108,** 1906–1923 (1994).
25. Sicilia-Aguilar, A. *et al.* The rapid outbursting star GM Cep: an EXor in Tr 37? *Astrophys. J.* **673,** 382–399 (2008).
26. Watson, D. M. *et al.* Crystalline silicates and dust processing in the protoplanetary disks of the Taurus young cluster. *Astrophys. J. Suppl. Ser.* **180,** 84–101 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. F. Jones for providing us with timely visual observations of EX Lupi during the preparation of our infrared measurements. The presented work was partly supported by the Hungarian Research Fund. The research of Á.K. is supported by the Netherlands Organization for Scientific Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.Á. (abraham@konkoly.hu).

LETTERS

Radiation-pressure mixing of large dust grains in protoplanetary disks

Dejan Vinković[1]

Dusty disks around young stars are formed out of interstellar dust that consists of amorphous, submicrometre grains. Yet the grains found in comets[1] and meteorites[2], and traced in the spectra of young stars[3], include large crystalline grains that must have undergone annealing or condensation at temperatures in excess of 1,000 K, even though they are mixed with surrounding material that never experienced temperatures as high as that[4]. This prompted theories of large-scale mixing capable of transporting thermally altered grains from the inner, hot part of accretion disks to outer, colder disk regions[5–7], but all have assumptions that may be problematic[8–12]. Here I report that infrared radiation arising from the dusty disk can loft grains bigger than one micrometre out of the inner disk, whereupon they are pushed outwards by stellar radiation pressure while gliding above the disk. Grains re-enter the disk at radii where it is too cold to produce sufficient infrared radiation-pressure support for a given grain size and solid density. Properties of the observed disks suggest that this process might be active in almost all young stellar objects and young brown dwarfs.

The history of thermal and compositional alternation of dust in dense, dusty protoplanetary disks around young pre-main-sequence stars enables us to better understand conditions that initiate formation of planets. One of the long-standing problems arising from this approach is the presence of crystalline dust in disk environments considered too cold for crystallinity to occur. Thus, it has been suggested that silicates crystallize in the hot part of the disk close to the central star and are transported outward into a colder environment. Currently favoured theories of outward transport include (1) turbulent mixing[6], (2) ballistic launching of particles in a dense wind created by interaction of the accretion disk with the young star's magnetic field (X-wind model)[5], and (3) mixing mediated by transient spiral arms in marginally gravitationally unstable disks[7]. Although these theories sound promising and may eventually result in the definitive solution to the problem of a large-scale mixing, they are so far hampered by theoretical assumptions needed for them to work. The turbulent mixing requires a source of efficient turbulent viscosity and the magnetorotational instability is invoked as the most promising candidate, but large stretches of the disk are considered not sufficiently ionized to keep this instability active[8–10]. The X-wind model relies on the theoretical notion of magnetic field configurations in the immediate vicinity of pre-main-sequence stars and high hopes are put on future observations to resolve this predicament[11]. The spiral arms model is in the domain of discussions on whether the underlying numerics, physical approximations and assumptions about the initial conditions are realistic enough to make results plausible[8,9,12].

Unlike these theories, non-radial radiation pressure does not require additional assumptions about the physical conditions in the disk, because it stems from the basic radiative transfer properties of optically thick dusty disks. It has been already shown that individual submicrometre grains do not move far away in the disk when pushed by radiation pressure because the force is primarily produced by radial stellar flux[13]. On the other hand, micrometre or larger grains are big enough to also have efficient interaction with the near-infrared photons (equivalent to dust temperatures of ~1,000–2,000 K) from the hot inner disk. Submicrometre grains are very inefficient emitters in the near-infrared; hence, they overheat and sublimate further away from the inner disk surface. This leaves the surface populated only with large grains, whereas small grains can survive within the optically thick interior[14] or at larger disk radii. Direct imaging with near-infrared interferometers has revealed that the observed location of the inner disk rim is consistent with this description (see refs 15–17 and references therein).

In optically thick protoplanetary disks, dust particles $\lesssim$1 mm are well coupled with the gas and their dynamics are dominated by the gas drag[13,18]. Hence, dust motion is very similar to the gas orbital, almost Keplerian, motion. Radiation-pressure force serves as a slow perturbation that leads to the rearrangement of dust orbits. To reconstruct the trajectory of particles pushed by radiation, we need to derive the spatial orientation of the radiation-pressure vector. For that, we need estimates of the diffuse flux as the source of pressure asymmetry. I solve this using the two-layer formalism, which is a well established method used in problems involving protoplanetary disk emission[19].

A short simplified solution is presented in Fig. 1, and a more rigorous derivation, which includes gravity, gas drag and radiation pressure, is described in Supplementary Information section 1. The result shows that the net radiation-pressure force, which combines stellar and diffuse flux components, is directed exactly parallel to the disk surface irrespective of its curvature. This leads to a very interesting scenario. If the force is strong enough to move a large dust grain, then such a crystalline grain formed at the hot inner rim would glide over the disk surface towards colder disk regions until the diffuse disk flux becomes too 'cold' (that is, its peak wavelength is larger than the dust size), at which point the force keeping the dust afloat ceases.

Further insight into the nature of non-radial radiation-pressure outflow requires a more detailed description of the disk structure and an advanced radiative transfer calculation. I start with preliminary modelling. The first results are presented in Fig. 2. The model assumes dusty disk density structure of the form $\rho_d(R, z) \propto R^{-2} \exp(-z^2/2h^2)$, with the scale height $h = 1.67 \times 10^{-2} R^{1.25}$, where R and z are cylindrical coordinates scaled with the dust sublimation radius R_{in} (the disk's inner rim; see Fig. 1), and ρ_d is the density of the dust fraction of the disk mass. The disk contains 0.1-μm and 2-μm olivine grains; the ratio of densities of particles in the disk is 10^4:1 and there is an overall radial visual optical depth at $z = 0$ of 10,000. I performed a full two-dimensional radiative transfer for the case of disk heating from a 10,000 K star, and dust sublimation at 1,500 K.

[1]Physics Department, University of Split, Nikole Tesle 12, 21000 Split, Croatia.

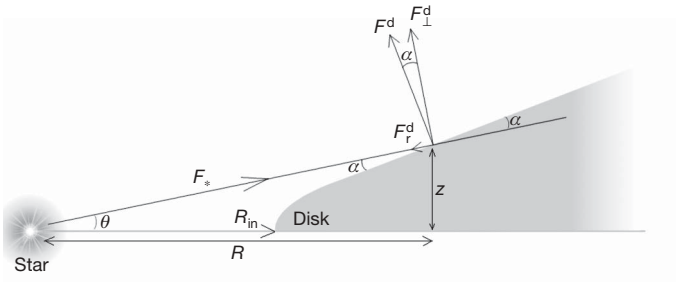


Figure 1 | Geometry of non-radial radiation pressure. The sketch shows a cross-section of an optically thick protoplanetary dusty disk heated by a star. The disk has a central hole of radius R_{in} where the dust overheats and sublimates away. According to the two-layer formalism[19], analysis of the disk emission in the near-infrared can be reduced to the disk's optically thin surface, which is heated directly by the stellar radiation. In this approach, the disk surface is replaced with a single temperature layer and we assume that the stellar radiation is completely absorbed within this layer. The disk interior is described as the second temperature layer, but it is heated only by infrared radiation from the surface layer and, therefore, is much colder and does not contribute to the disk emission in the near- and mid-infrared[19]. In optically thick passive disks we can use energy conservation at a surface point (R, z) to set balance, $F_* \sin\alpha = F^d$, between the bolometric stellar flux F_* intercepted by the disk at a grazing angle α and the outgoing disk radiation F^d (infrared emission and scattered stellar photons). In the approximation of a geometrically thin disk surface, we can assume that the entire local diffuse flux at the very surface is perpendicular to the surface. Grains that manage to decouple and move away from the surface would feel a reduced flux as the diffuse radiation streams out in all directions. We can decompose F^d into radial, $F^d_r = -F_* \sin^2\alpha$, and azimuthal, $F^d_\perp = F_* \sin\alpha \cos\alpha$, components. If dust grains are big enough to have constant extinction in the wavelength range of F^d_λ, then the radiation-pressure force becomes $F \propto F_* + F_d$. Using flux components from above gives the radial force $F_r \propto F_* \cos^2\alpha$ and the azimuthal force $F_\perp \propto F_* \sin\alpha \cos\alpha$. Notice that this yields radiation-pressure force directed exactly parallel to the disk surface, $F_\perp/F_r = \tan\alpha$, irrespective of the disk curvature. A more rigorous derivation is presented in Supplementary Information section 1, including dust dynamics due to gravity, gas drag and radiation pressure.

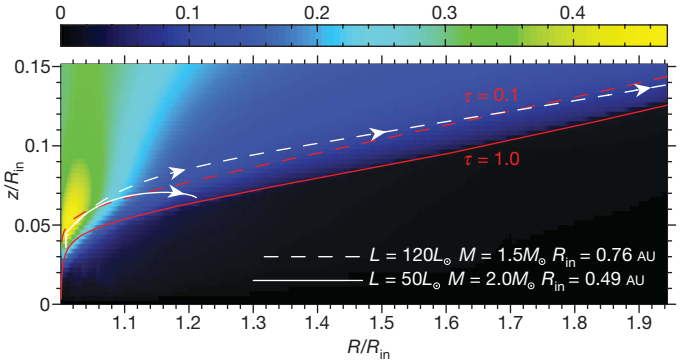


Figure 2 | Trajectory of dust grains under the influence of stellar gravity, gas drag and non-radial radiation pressure. The coloured background map shows the vertical z component of the radiation-pressure vector scaled with the value that the stellar pressure would have if the dusty disk were not there. Two stellar luminosity/mass ratios are used: $80L_\odot/M_\odot$ (white dashed line) and $25L_\odot/M_\odot$ (white solid line). Dust composition is olivine[29] of 2 μm radius and solid density 3,000 kg m^{-3}. Dust grains start their travel with a vertical upward motion until the gas density drops enough to loosen the influence of gas drag. After that the grain is ejected to a larger disk radius, where trajectory details depend on the strength and direction of radiation pressure. Trajectory is calculated numerically with the Runge–Kutta method. Radiation pressure is calculated numerically from two-dimensional radiative transfer that includes dust absorption, scattering and emission. The disk consists of 0.1-μm and 2-μm grains that sublimate at 1,500 K, but the surface in this disk region is too hot for 0.1-μm grains, which survive below the surface populated by 2-μm grains. Spatial dimensions are scaled with the disk sublimation radius, R_{in} (see Fig. 1). The disk gas and dust densities decrease exponentially with z. Red lines show the disk surfaces defined by the radial visual optical depth $\tau = 0.1$ (dashed red line) and $\tau = 1$ (solid red line). Details of this numerical result will be given elsewhere.

Location of the dust sublimation disk surface is calculated self-consistently from the mutual exchange of infrared energy between 0.1-μm and 2-μm grains, resulting in $R_{in} = 44.7R_*$ (R_* is the stellar radius). Figure 2 shows the map of vertical radiation pressure on 2-μm grains and examples of grain trajectories. Results from this detailed approach confirm the plausibility of my theoretical arguments.

The ability of large grains to migrate along any disk curvature makes this theory independent of the ongoing debate about the geometrical structure of the inner disk region[15]. The popular view is that the inner sublimation edge is puffed-up and curved[17]. The non-radial pressure would affect dust dynamics under such a disk curvature in the same way as in the numerical example above, except that individual grains could decouple more easily from the inner disk and fly towards outer disk regions owing to the disk's self-shadowing[20].

Grains pushed by radiation create an outflow that operates at much shorter timescales than the local dust settling, because radiation pressure is active in the region of lower gas density. In Supplementary Information section 2, I provide an estimate of the total amount of dust that flows outward in the disk surface layer. The outflow strength and range depend on the ratio, β, of radiation force tangential to the disk surface to the local gravity force (see Supplementary Information section 1 for a detailed description):

$$\beta \approx 0.4 L_* (M_* \rho_s a)^{-1} \quad (1)$$

where L_* is the stellar luminosity (in units of solar luminosity, $L_\odot$), M_* is the stellar mass (in units of solar mass, $M_\odot$), ρ_s is the grain solid density (in units of 3,000 kg m^{-3}), and a is the dust grain radius (in micrometres). Grains with $\beta \gtrsim 0.5$ are gravitationally decoupled from the star, and will be pushed away from the star as long as the diffuse flux keeps them afloat within the optically thin surface. Grains with $\beta \lesssim 0.5$ feel a 'reduced' gravity and their settling is slowed down.

I attempted to estimate the spatial extent of significant vertical radiation pressure along the disk surface. I used a simplified, but illustrative model of the protoplanetary disk where the disk surface contains only single-size grains. Results show (see Fig. 3 and Supplementary Information section 3) that significant dynamical effects from the non-radial radiation pressure are possible only for grains bigger than about 1 μm. Grains a few micrometres in size can be lifted out of the disk only at small disk radii where the disk is hottest, but already 5-μm grains can 'glide' to large radii (over 1,000 stellar radii), provided that the radiation pressure is strong enough to push such a grain. The upper limit on grain size pushed that way is dictated by equation (1), which shows how the force decreases with grain size.

Notice that I assume a solid spherical grain, which is a simplification of a more realistic fluffy dust aggregate[21,22]. Aggregates result in a much larger β for the same grain size because they have a much lower grain density than the typical 3,000 kg m^{-3} owing to inclusion of vacuum into the grain structure. On the other hand, crystalline grains are largely transparent in the spectral range of stellar radiation[1], which would make radiation pressure ineffective. This remains an open problem for my theory, although crystalline grains incorporated into dust aggregates might have a non-transparent 'glue' keeping the aggregate together, which would increase β and mitigate these problems. Such 'dustballs' are considered to be precursors of chondrules and CAIs (calcium–aluminium inclusions) in meteorites[23].

The main stellar parameter dictating the overall strength β of the radiation-pressure effect on a grain is the luminosity/mass ratio, L_*/M_*. Observations and evolutionary tracks indicate that $L_* = M_* \gtrsim 0.5$ (which gives $\beta \approx 0.4$ for a grain of 1-μm diameter) in almost all young stellar objects, including brown dwarfs. Thus, non-radial radiation pressure is at least marginally relevant in all these objects, especially if a realistic dust aggregate model is taken

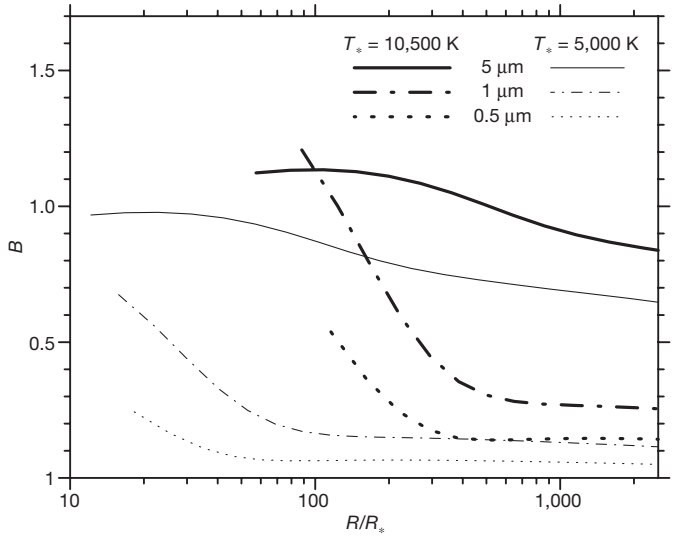


Figure 3 | Estimated strength of diffuse radiation pressure along the disk surface, indicating how far grains can travel. The estimated strength, B, is defined as the ratio of diffuse to stellar β perpendicular to the disk surface (see equation (34) in Supplementary Information section 3 for details), at various distances from the star. Optical properties of the pushed grains and dust forming the disk surface are the same. The surface contains only one grain size and type. Lines show results for spherical grains of 0.5-μm, 1-μm and 5-μm radius for two different stellar temperatures T_*. Diffuse radiation pressure is important (that is, $B \approx 1$) only for grains $\gtrsim 1$ μm. Grains $\gtrsim 5$ μm experience strong diffuse pressure over a large disk surface because of their efficient infrared absorption at longer wavelengths, whereas smaller micrometre grains can float only above the inner disk with the highest temperature. The dust is enstatite[30]. Other compositions lead to qualitatively similar curves. Lines start at radii defined by the 1,500 K dust sublimation temperature.

into account. Moreover, at earlier evolutionary stages β would have been larger because, according to stellar evolution models, the end of significant accretion (99% of the final mass) occurs with $L_*/M_* > 10$ for stars with $M_* \lesssim 1 M_\odot$ (ref. 24).

As crystallization is very efficient along the hot inner disk rim, radiation-pressure mixing of large grains would inevitably include the crystalline fraction and disperse such dust over the disk surface. Interestingly, such a correlation between large grains and crystalline fraction is detected in Herbig Ae stars (see, for example, refs 25, 26). This would be most pronounced in the inner disk regions, closer to the inner rim, as is indeed observed (see, for example, refs 3, 27, 28). With the help of disk turbulence, the surface of the inner disk region is constantly replenished with new grains and the process continues as long as the radiation pressure is active.

Received 11 January; accepted 31 March 2009.

1. Brownlee, D. *et al.* Comet 81P/Wild 2 under a microscope. *Science* **314,** 1711–1716 (2006).
2. Wooden, D. H., Harker, D. E. & Brearley, A. J. in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) 774–808 (ASP Conference Series Vol. 341, 2005).
3. van Boekel, R. *et al.* The building blocks of planets within the 'terrestrial' region of protoplanetary disks. *Nature* **432,** 479–482 (2004).
4. Hill, H. G. M., Grady, C. A., Nuth, J. A. III, Hallenbeck, S. L. & Sitko, M. L. Constraints on nebular dynamics and chemistry based on observations of annealed magnesium silicate grains in comets and in disks surrounding Herbig Ae/Be stars. *Proc. Natl Acad. Sci. USA* **98,** 2182–2187 (2001).
5. Shu, F. H., Shang, H. & Lee, T. Toward an astrophysical theory of chondrites. *Science* **271,** 1545–1552 (1996).
6. Ciesla, F. J. Outward transport of high-temperature materials around the midplane of the solar nebula. *Science* **318,** 613–615 (2007).
7. Boss, A. P. Mixing in the solar nebula: Implications for isotopic heterogeneity and large-scale transport of refractory grains. *Earth Planet. Sci. Lett.* **268,** 102–109 (2008).
8. Klahr, H., Różyczka, M., Dziourkevitch, N., Wünsch, R. & Johansen, A. in *Planet Formation* (eds Klahr, H. & Brandner, W.) 42–63 (Cambridge Univ. Press, 2006).
9. Hubickyj, O. in *Planet Formation* (eds Klahr, H. & Brandner, W.) 163–178 (Cambridge Univ. Press, 2006).
10. Matsumura, S. & Pudritz, R. E. Dead zones and extrasolar planetary properties. *Mon. Not. R. Astron. Soc.* **365,** 572–584 (2006).
11. Shang, H., Li, Z.-Y. & Hirano, N. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 261–276 (Univ. Arizona Press, 2007).
12. Boss, A. P. Evolution of the solar nebula. VII. Formation and survival of protoplanets formed by disk instability. *Astrophys. J.* **629,** 535–548 (2005).
13. Takeuchi, T. & Lin, D. N. C. Surface outflow in optically thick dust disks by radiation pressure. *Astrophys. J.* **593,** 524–533 (2003).
14. Vinković, D. Temperature inversion on the surface of externally heated optically thick multigrain dust clouds. *Astrophys. J.* **651,** 906–913 (2006).
15. Millan-Gabet, R. *et al.* in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 539–554 (Univ. Arizona Press, 2007).
16. Vinković, D. & Jurkić, T. Relation between the luminosity of young stellar objects and their circumstellar environment. *Astrophys. J.* **658,** 462–479 (2007).
17. Isella, A. & Natta, A. The shape of the inner rim in proto-planetary disks. *Astron. Astrophys.* **438,** 899–907 (2005).
18. Alexander, R. From discs to planetesimals: Evolution of gas and dust discs. *N. Astron. Rev.* **52,** 60–77 (2008).
19. Chiang, E. I. & Goldreich, P. Spectral energy distributions of T Tauri stars with passive circumstellar disks. *Astrophys. J.* **490,** 368–376 (1997).
20. Fujiwara, D. & Watanabe, S. *Protostars and Planets V, Proceedings of the Conference* 8237 (Lunar and Planetary Institute, 2005).
21. Voshchinnikov, N. V. & Henning, T. Is the silicate emission feature only influenced by grain size? *Astron. Astrophys.* **483,** L9–L12 (2008).
22. Pinte, C. *et al.* Probing dust grain evolution in IM Lupi's circumstellar disc. Multi-wavelength observations and modelling of the dust disc. *Astron. Astrophys.* **489,** 633–650 (2008).
23. Jones, R. H., Lee, T., Connolly, H. C. Jr, Love, S. G. & Shang, H. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 927–962 (Univ. Arizona Press, 2000).
24. Wuchterl, G. & Tscharnuter, W. M. From clouds to stars. Protostellar collapse and the evolution to the pre-main sequence I. Equations and evolution in the Hertzsprung-Russell diagram. *Astron. Astrophys.* **398,** 1081–1090 (2003).
25. van Boekel, R. *et al.* A 10 μm spectroscopic survey of Herbig Ae star disks: Grain growth and crystallization. *Astron. Astrophys.* **437,** 189–208 (2005).
26. Sargent, B. *et al.* Dust processing in disks around T Tauri stars. *Astrophys. J.* **645,** 395–415 (2006).
27. Watson, D. M. *et al.* Crystalline silicates and dust processing in the protoplanetary disks of the Taurus young cluster. *Astrophys. J. Suppl. Ser.* **180,** 84–101 (2009).
28. Schegerer, A. A., Wolf, S., Ratzka, Th. & Leinert, Ch. The T Tauri star RY Tauri as a case study of the inner regions of circumstellar dust disks. *Astron. Astrophys.* **478,** 779–793 (2008).
29. Dorschner, J., Begemann, B., Henning, T., Jaeger, C. & Mutschke, H. Steps toward interstellar silicate mineralogy. II. Study of Mg-Fe-silicate glasses of variable composition. *Astron. Astrophys.* **300,** 503–520 (1995).
30. Jaeger, C. *et al.* Steps toward interstellar silicate mineralogy. IV. The crystalline revolution. *Astron. Astrophys.* **339,** 904–916 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank the Institute for Advanced Study in Princeton and the University Computing Center SRCE in Zagreb for time on their computer clusters.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to the author (vinkovic@pmfst.hr).

Vol 459|14 May 2009|doi:10.1038/nature07998

LETTERS

Thermal vestige of the zero-temperature jamming transition

Zexin Zhang[1]*, Ning Xu[1,2]*, Daniel T. N. Chen[1], Peter Yunker[1], Ahmed M. Alsayed[1], Kevin B. Aptowicz[3], Piotr Habdas[4], Andrea J. Liu[1], Sidney R. Nagel[2] & Arjun G. Yodh[1]

When the packing fraction is increased sufficiently, loose particulates jam to form a rigid solid in which the constituents are no longer free to move. In typical granular materials and foams, the thermal energy is too small to produce structural rearrangements. In this zero-temperature ($T=0$) limit, multiple diverging[1–8] and vanishing[2,9,10] length scales characterize the approach to a sharp jamming transition. However, because thermal motion becomes relevant when the particles are small enough, it is imperative to understand how these length scales evolve as the temperature is increased. Here we used both colloidal experiments and computer simulations to progress beyond the zero-temperature limit to track one of the key parameters—the overlap distance between neighbouring particles—which vanishes at the $T=0$ jamming transition. We find that this structural feature retains a vestige of its $T=0$ behaviour and evolves in an unusual manner, which has masked its appearance until now. It is evident as a function of packing fraction at fixed temperature, but not as a function of temperature at fixed packing fraction or pressure. Our results conclusively demonstrate that length scales associated with the $T=0$ jamming transition persist in thermal systems, not only in simulations but also in laboratory experiments.

The onset of the arrested dynamics associated with jamming depends on an interplay between packing constraints, thermal energy and applied forcing[11,12]. This behaviour is illustrated in the schematic jamming phase diagram of Fig. 1, where the zero-temperature jamming transition point for finite-range, repulsive spheres[2] is labelled 'J'. It has been unclear how this zero-temperature transition affects behaviour at non-zero temperature. To explore its influence, we used experiments and numerical simulations to study structure and dynamics at non-zero temperature in the vicinity of Point J.

At zero temperature, the average number of touching neighbours per particle, Z, jumps discontinuously at Point J, from $Z=0$ to the minimum number required for mechanical stability, Z_c, when the packing fraction, ϕ, is increased through the transition at ϕ_c (refs 1, 2 and 13). This discontinuity produces a δ-function in the first peak of the pair-correlation function $g(r)$, which measures the probability of finding another particle at distance r given one at the origin[2,10]. Numerical simulations at $T=0$ confirm that g_1, the height of the first peak in $g(r)$, diverges as $g_1 \approx |\phi-\phi_c|^{-1.0}$ as ϕ_c is approached both from above[2,10] and below[9]. The overlap distance L_{ov} (that is, the left-hand width of the first peak) is directly related to g_1 because $g_1 L_{ov} \approx Z_c$ near the transition. Thus, a maximum in g_1 corresponds to a minimum in L_{ov}, and the divergence in g_1 at the transition corresponds to the vanishing of the overlap distance[9] as $L_{ov} \approx |\varphi-\varphi_c|^{1.0}$. Here we explored how the overlap distance, as measured by the height of the first peak of $g(r)$, evolves as a function of temperature near Point J.

In two-dimensional colloidal experiments, we probed the jamming transition at non-zero temperature by tuning the packing fraction. The experimental trajectory closely follows a horizontal line at fixed temperature in the $T-(1/\phi)$ plane above Point J in the jamming phase diagram (Fig. 1, dashed line). In parallel, we used three-dimensional simulations to explore the jamming transition in the same $T-(1/\phi)$ plane by two routes: (1) varying the packing fraction at fixed temperature (Fig. 1, dashed line) and (2) varying the temperature at fixed pressure (Fig. 1, dotted line).

The colloidal samples were aqueous suspensions of poly(*N*-isopropyl acrylamide) microgel colloidal particles (NIPA particles)[14,15], whose diameters increase substantially as temperature is reduced only slightly. Therefore, sample packing fraction could be tuned over a wide

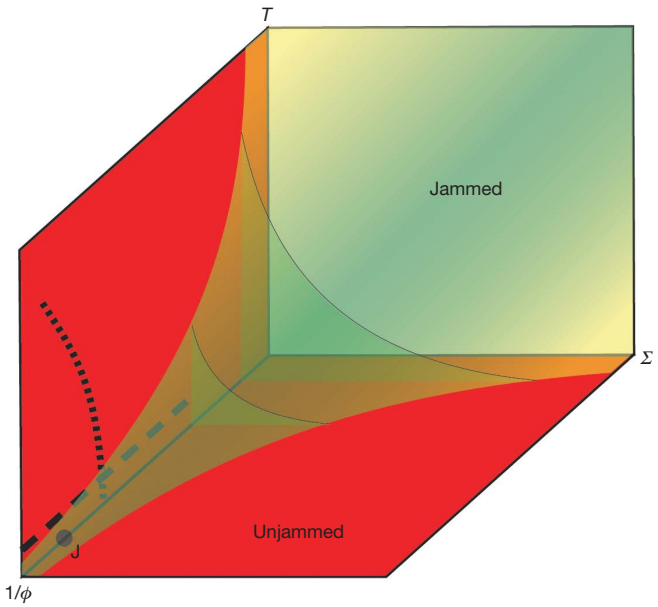


Figure 1 | Schematic jamming phase diagram. The surface of the green region in the three-dimensional space defined by temperature T, inverse packing fraction $1/\phi$ and applied stress Σ corresponds to the dynamical glass transition; within the green region the system is out of equilibrium. The point marked J represents a phase transition that occurs as ϕ is increased while $T=0$ and $\Sigma=0$. In the experiments, we varied the packing fraction at nearly fixed temperature, along the horizontal dashed line. In the simulations, we vary both packing fraction at fixed temperature along the horizontal dashed line, and temperature at fixed pressure along the dotted curve.

[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [2]James Franck Institute, University of Chicago, Chicago, Illinois 60637, USA. [3]Department of Physics, West Chester University, West Chester, Pennsylvania 19383, USA. [4]Department of Physics, Saint Joseph's University, Philadelphia, Pennsylvania 19131, USA.
*These authors contributed equally to this work.

range with only minimal changes of temperature. This class of suspension has been successfully used to model a variety of phase transitions[16–21]. In our experiments, approximately equal numbers of monodisperse small and large NIPA particles with room-temperature (25 °C) diameters of $\sigma_S = 1.17$ μm and $\sigma_L = 1.63$ μm, respectively, were sandwiched between two glass cover slips to form a two-dimensional colloidal suspension. The particle interaction potentials were measured to be short-range repulsive with a soft tail (Supplementary Fig. S2). The use of binary mixtures reduces the possibility of crystallization[22,23] and the softness of the potential, in contrast to that of hard spheres, permits access to packing fractions above the jamming transition.

In most colloidal experiments the thermodynamic control variable is packing fraction. Temperature control elements on the microscope objective in our experiments permitted the packing fraction ϕ to be varied *in situ* from ~0.76 to ~0.93, that is, across the packing fraction of the $T = 0$ jamming transition at $\phi \approx 0.84$ for temperatures ranging from 24.0 °C to 30.0 °C. At each ϕ the sample was permitted to equilibrate for 1,200 s before measurements were taken. We then used standard video microscopy[24] and particle-tracking techniques[25] to obtain the particle positions and the particle displacements. By identifying particle size and position we computed the three distinct pair-correlation functions: g_{LL} associated with large particles only, g_{SS} associated with small particles only, and g_{LS} probing the correlation between large and small particles. Here we focus only on g_{LL}. Qualitatively similar results were obtained for the other two correlation functions (Supplementary Fig. S3).

Figure 2 shows g_{LL} as a function of packing fraction ϕ. A prominent first peak at a distance of approximately one large particle diameter was found at all ϕ. In the inset to Fig. 2 we plotted g_1, the height of the first peak of $g_{LL}(r)$, versus ϕ. We note that g_1 has a pronounced maximum at $\phi = 0.85$. We identify this maximum as a vestige of the divergence in $g(r)$ seen at Point J, the $T = 0$ jamming transition.

In parallel, we used molecular dynamics simulations to explore the maximum in g_1 as a function of T and ϕ. We performed simulations using 1,000 particles of mass m in a three-dimensional cubic box with periodic boundary conditions. The particles are taken from a 50:50 distribution of the two diameters σ_L and σ_S, with ratio $\sigma_L/\sigma_S = 1.4$. Particles i and j interact via a repulsive spring-like potential, $U(r_{ij}) = \varepsilon\left(1 - r_{ij}/d_{ij}\right)^{\alpha}/\alpha$, if their separation r_{ij} is smaller than the sum of their radii (that is, if they overlap), and do not interact otherwise. We used two types of repulsive potentials: harmonic ($\alpha = 2$) and Hertzian ($\alpha = 5/2$). We express distance in units of σ_S, time in units of $\sqrt{m\sigma_S^2/\varepsilon}$, sample temperature T in units of ε and pressure in units of ε/σ_S^3. We note that the Hertzian form provides a reasonable fit to the experimentally measured pair potential for NIPA particles at low concentration, with $\varepsilon/T \cong 270$ for the large particles (Supplementary Fig. S2). Figure 3a shows the data from simulations for harmonic potentials at four temperatures that are analogous to our ϕ-dependent colloid experiments: g_1 is plotted versus $\Delta\phi \equiv \phi - \phi_c$, where ϕ_c is the onset of jamming at $T = 0$. The curve for each T exhibits a clear maximum, where $g_1 = g_1^{max}$, at $\Delta\phi_v(T)$ (subscript 'v' indicates vestige; inset to Fig. 3a). Thus, the constant-temperature three-dimensional simulation data are consistent with the colloidal experiments in two dimensions in that they both exhibit structural maxima as a function of packing fraction.

In both simulation and experiment, the value of g_1^{max} is finite and does not diverge as it does at Point J (refs 2, 9). In experiments, many factors can conspire to reduce g_1^{max}. In simulations, however, g_1^{max} is finite only because the temperature is not zero. Indeed, Fig. 3a shows that g_1^{max} decreases with increasing T as $g_1^{max} \propto (\Delta\varphi_v(T))^{-1}$, while its inset shows that $\Delta\phi_v(T)$ approaches zero as T tends to zero. This behaviour demonstrates that the maximum in g_1 at non-zero T evolves directly from the divergence in g_1^{max} at Point J.

The existence of a maximum in g_1 at finite temperature is easily understood. In the dilute limit, the height of the first peak increases with ϕ as more particles join the first-neighbour shell. At high ϕ, the first peak broadens with ϕ as the particles have greater overlap, leading to a drop in the peak height. We can predict the ϕ dependence of g_1^{max} as follows. At finite temperature, there are two contributions to the overlap between particles: (1) the static overlap L_O due to compression, which would exist even at $T = 0$, and (2) the additional overlap L_T

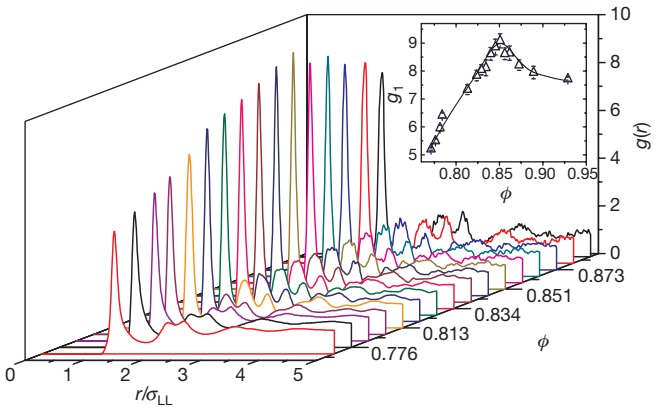


Figure 2 | Pair-correlation function *g(r)* for the large particles at all experimental packing fractions. The inset shows g_1, the height of the first peak of $g(r)$, as a function of packing fraction ϕ. The error bars in g_1 are the standard deviations of three independent calculations.

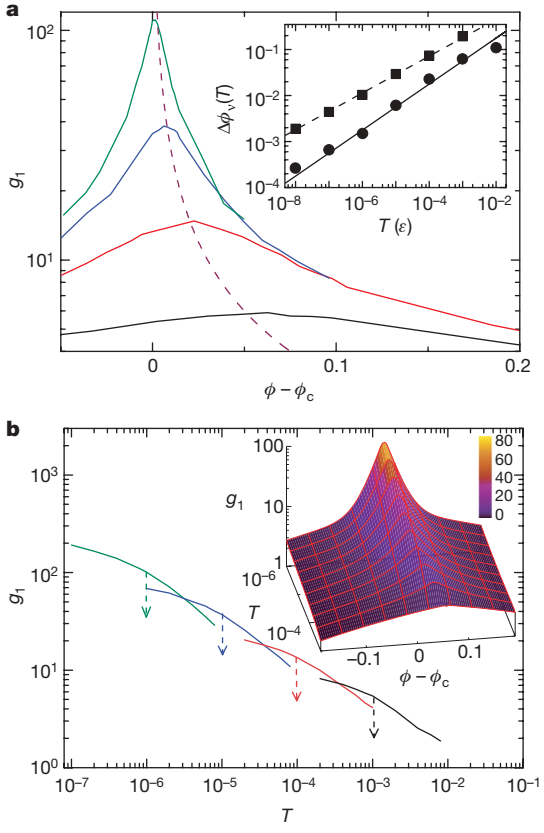


Figure 3 | Peak value of *g(r)*, g_1, measured from simulations. **a**, g_1 versus the packing fraction $\phi - \phi_c$, measured at temperatures $T = 10^{-3}\varepsilon$ (black), $10^{-4}\varepsilon$ (red), $10^{-5}\varepsilon$ (blue), and $10^{-6}\varepsilon$ (green), for harmonic interactions. The dashed line represents $g_1^{max} \propto \Delta\phi^{-1}$, as expected theoretically. The inset shows $\Delta\phi_v(T)$, the location of the structural maximum for harmonic repulsions (circles) and Hertzian repulsions (squares). The solid and dashed lines are fits to the expected power-law scaling: $\Delta\phi_v \propto T^{1/\alpha}$ for $\alpha = 2$ and 5/2, respectively. **b**, g_1 versus T measured at constant pressures $P = 0.023\varepsilon/\sigma_S^3$ (black), $0.067\varepsilon/\sigma_S^3$ (red), $0.0017\varepsilon/\sigma_S^3$ (blue), and $0.00067\varepsilon/\sigma_S^3$ (green) with the arrows pointing to the temperatures at which g_1 reaches the maximum measured by varying the packing fraction at constant T as shown in **a**. The inset shows a three-dimensional plot of g_1 (colour scale) versus T and $\phi - \phi_c$.

due to collisions arising from the (thermal) kinetic energy. The maximum in g_1 occurs when the spread in distances between neighbours is a minimum, which is typically when the total overlap, $L_{ov} = L_O + L_T$, is smallest. For sufficiently small T, the average potential energy per contact can be expanded as: $U(L_O + L_T) \cong U(L_O) + L_T \left.\frac{dU}{dL}\right|_{L=L_O}$. The system exhibits harmonic fluctuations around the energy minimum $U(L_O)$, so we have by equipartition that $U(L_O + L_T) - U(L_O) \propto T$. Note[2] that for a repulsive potential of the form $U(L) = \varepsilon L^\alpha$, we have $\left.\frac{dU}{dL}\right|_{L=L_O} \propto \Delta\phi^{\alpha-1}$ and $L_O \propto \Delta\varphi$ for sufficiently small $\Delta\phi$. Minimizing L_{ov} with respect to $\Delta\phi$ at fixed T therefore yields $\Delta\phi_v \propto T^{1/\alpha}$. The inset to Fig. 3a shows that this scaling is indeed observed in the simulations, confirming the view that the maximum in g_1 is a thermal structural vestige of the $T = 0$ jamming transition.

Although direct measurements of the pair-correlation function in three-dimensional colloidal systems have been made on colloidal glasses[26,27], to our knowledge the structural feature presented above has not been observed. A maximum in g_1 was observed in an athermal gas-fluidized granular system with increasing density at non-zero kinetic energy[28], with a second rise at the approach to random close-packing at zero kinetic energy. It is possible that the kinetic-energy/density trajectory of that experiment intersects the curve marking the evolution of the structural vestige with kinetic energy (or temperature) twice, once at the first local maximum and once at a second local maximum at Point J.

A maximum in g_1 was also not observed in scattering experiments on glass-forming liquids[29]. Such experiments extract positional information via measurements of the Fourier transform of the pair correlation function, the structure factor. One can readily show that a sharpening, or even a divergence of the first peak in $g(r)$, transforms into a signature in the structure factor that is spread over a wide range of wavevectors and is too subtle to be resolved with realistic experimental signal-to-noise conditions[2].

However, many simulations have searched for structural signatures of the glass transition in $g(r)$ (refs 22, 30). How could these simulations not have seen a maximum in the height of the first peak of $g(r)$? To answer this question, we conducted simulations along the more traditional phase-space trajectory, applicable to supercooled liquids and glasses, wherein temperature is varied as pressure (or packing fraction) is kept constant. Figure 3b shows that g_1 increases monotonically and does not exhibit a maximum when T is lowered at fixed pressure. Therefore, we do not see the structural vestige of Point J in a typical trajectory used to study the glass transition; we see a feature only when packing fraction or pressure is varied at fixed temperature. The behaviour of g_1 as a function of both T and ϕ is shown in the inset to Fig. 3b. Our observations are thus consistent with previous simulations, none of which explored trajectories at fixed temperature.

The systems studied here also exhibit classic dynamical glass transitions in which the structural relaxation time reaches the maximum timescale of the experiment or simulation. The dynamical glass transition lies at the boundary between the jammed and unjammed regions in the $T - (1/\phi)$ plane shown in the jamming phase diagram (Fig. 1). It is important to understand where the structural vestige of Point J, identified here from the structural maximum, lies in relation to the glass-transition line. To locate the dynamical glass transition in both the experiment and simulation, we measured the relaxation time τ_α, determined from the coherent intermediate scattering function (defined in the Supplementary Information). Experimentally, we found that τ_α increases rapidly with ϕ and eventually surpasses the experimental time window at $\phi_g \approx 0.85$, thus defining the packing fraction of the dynamical glass transition (Fig. 4a). This packing fraction coincides with the location of the maximum of g_1. Thus, in the colloidal experiment the thermal vestige of Point J occurs near the same packing fraction $\phi_v \cong 0.85$ as the dynamical glass transition so that $\phi_v \cong \phi_g$. However, this is not the case for the simulations, which

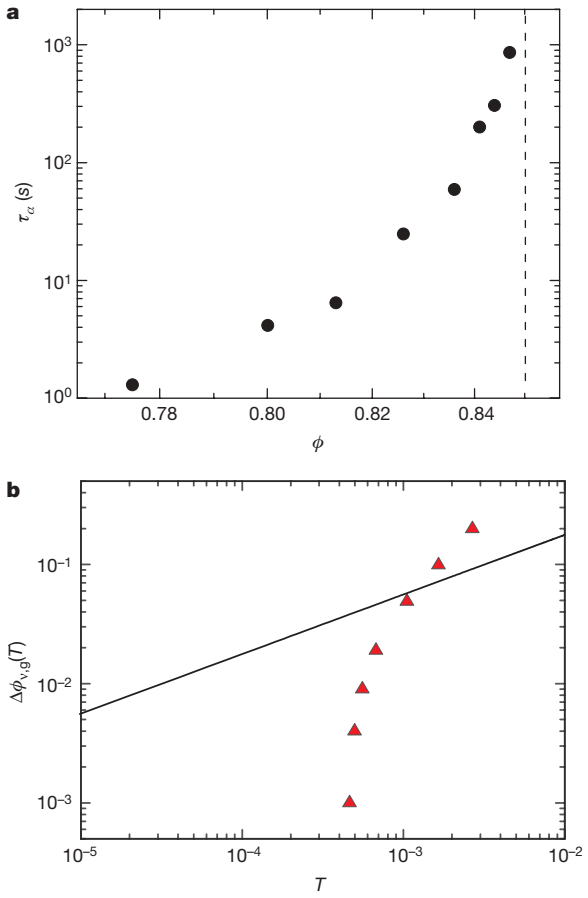


Figure 4 | Dynamics approaching the structural maximum. a, Experimental results for the α relaxation time τ_α for several packing fractions ϕ. The vertical dashed red line denotes the location of the structural maximum, determined from Fig. 2. **b**, Simulation results for $\Delta\phi_g(T)$ (red solid triangles), defined by where the relaxation time is equal to 10^4 in units of $\sqrt{m\sigma_S^2/\varepsilon}$, and $\Delta\phi_v(T)$ (black solid line, reproduced from the inset to Fig. 3a, corresponding to the power-law fit), the location of the structural maximum. Both $\Delta\phi_g(T)$ and $\Delta\phi_v(T)$ are calculated for harmonic repulsions ($\alpha = 2$).

can measure both ϕ_v and ϕ_g as a function of temperature. Thus, the experimental observation that $\phi_v \cong \phi_g$ appears to be a coincidence.

To demonstrate this in our simulations, we find the packing fraction of the dynamical glass transition, $\Delta\phi_g \equiv \phi_g - \phi_c$ for each temperature T at which τ_α exceeds the measurable window. In Fig. 4b we compare $\Delta\phi_g(T)$ to $\Delta\phi_v(T) \equiv \phi_v(T) - \phi_c$, the location of the structural vestige of jamming transition. At low temperatures, $\Delta\phi_g(T) < \Delta\phi_v(T)$, whereas at higher temperatures, $\Delta\phi_g(T) > \Delta\phi_v(T)$. In cases where the jamming transition lies in the out-of-equilibrium regime of Fig. 1, we find that g_1 and $\Delta\phi_v(T)$ are very robust to sample history. Rapidly quenched samples do age slightly, but settle down to a value of g_1 consistent with the results for slow quenches. Thus the vestige is neither a structural signature of the glass transition nor an artefact of falling out of equilibrium.

To conclude, we studied jamming in thermal systems in the vicinity of Point J. We found a maximum in the height of the first peak of the pair-correlation function that shifts to higher packing fractions as the temperature is increased from zero as $\Delta\phi_v \propto T^{1/\alpha}$, where α characterizes the inter-particle potential. This maximum is a vestige of one of the most important length scales that define the zero-temperature jamming transition at Point J (that is, the overlap length L_{ov} between neighbours). At Point J, this length scale vanishes because the system is isostatic and on the brink of mechanical failure. The present work shows that the evolution of the jamming transition with temperature is now accessible to experimental attack in colloidal systems. For

example, the evolution of two other diverging lengths at Point J, derived from the density of vibrational states and the elastic moduli[2], could be followed by experiment because the density of normal modes of vibration can, in principle, be measured from the Fourier transformation of the displacement of an individual particle in a colloidal sample. These length scales hold the possibility of connection to the glass transition, given that diverging timescales are often associated with diverging length scales. Our observations therefore demonstrate that length scales associated with the $T = 0$ jamming transition persist at non-zero temperatures, and also provide a route for using colloids to explore the relationship between Point J and the glass transition.

Received 10 January; accepted 13 March 2009.

1. O'Hern, C. S., Langer, S. A., Liu, A. J. & Nagel, S. R. Random packings of frictionless particles. *Phys. Rev. Lett.* **88,** 075507 (2002).
2. O'Hern, C. S., Silbert, L. E., Liu, A. J. & Nagel, S. R. Jamming at zero temperature and zero applied stress: The epitome of disorder. *Phys. Rev. E* **68,** 011306 (2003).
3. Drocco, J. A., Hastings, M. B., Reichhardt, C. J. O. & Reichhardt, C. Multiscaling at point J: Jamming is a critical phenomenon. *Phys. Rev. Lett.* **95,** 088001 (2005).
4. Ellenbroek, W. G., Somfai, E., van Hecke, M. & van Saarloos, W. Critical scaling in linear response of frictionless granular packings near jamming. *Phys. Rev. Lett.* **97,** 258001 (2006).
5. Olsson, P. & Teitel, S. Critical scaling of shear viscosity at the jamming transition. *Phys. Rev. Lett.* **99,** 178001 (2007).
6. Silbert, L. E., Liu, A. J. & Nagel, S. R. Vibrations and diverging length scales near the unjamming transition. *Phys. Rev. Lett.* **95,** 098301 (2005).
7. Wyart, M., Silbert, L. E., Nagel, S. R. & Witten, T. A. Effects of compression on the vibrational modes of marginally jammed solids. *Phys. Rev. E* **72,** 051306 (2005).
8. Xu, N., Vitelli, V., Wyart, M., Liu, A. J. & Nagel, S. R. Energy transport in jammed sphere packings. *Phys. Rev. Lett.* **102,** 038001 (2009).
9. Silbert, L. E., Liu, A. J. & Nagel, S. R. Structural signatures of the unjamming transition at zero temperature. *Phys. Rev. E* **73,** 041304 (2006).
10. Donev, A., Torquato, S. & Stillinger, F. H. Pair correlation function characteristics of nearly jammed disordered and ordered hard-sphere packings. *Phys. Rev. E* **71,** 011105 (2005).
11. Liu, A. J. & Nagel, S. R. Nonlinear dynamics—Jamming is not just cool any more. *Nature* **396,** 21–22 (1998).
12. Trappe, V., Prasad, V., Cipelletti, L., Segre, P. N. & Weitz, D. A. Jamming phase diagram for attractive particles. *Nature* **411,** 772–775 (2001).
13. Durian, D. J. Foam mechanics at the bubble scale. *Phys. Rev. Lett.* **75,** 4780–4783 (1995).
14. Pelton, R. Temperature-sensitive aqueous microgels. *Adv. Colloid Interf. Sci.* **85,** 1–33 (2000).
15. Saunders, B. R. & Vincent, B. Microgel particles as model colloids: theory, properties and applications. *Adv. Colloid Interf. Sci.* **80,** 1–25 (1999).
16. Senff, H. & Richtering, W. Temperature sensitive microgel suspensions: Colloidal phase behavior and rheology of soft spheres. *J. Chem. Phys.* **111,** 1705–1711 (1999).
17. Wu, J. Z., Zhou, B. & Hu, Z. B. Phase behavior of thermally responsive microgel colloids. *Phys. Rev. Lett.* **90,** 048304 (2003).
18. Lyon, L. A. *et al.* Microgel colloidal crystals. *J. Phys. Chem. B* **108,** 19099–19108 (2004).
19. Alsayed, A. M., Islam, M. F., Zhang, J., Collings, P. J. & Yodh, A. G. Premelting at defects within bulk colloidal crystals. *Science* **309,** 1207–1210 (2005).
20. Han, Y., Ha, N. Y., Alsayed, A. M. & Yodh, A. G. Melting of two-dimensional tunable-diameter colloidal crystals. *Phys. Rev. E* **77,** 041406 (2008).
21. Han, Y. L. *et al.* Geometric frustration in buckled colloidal monolayers. *Nature* **456,** 898–903 (2008).
22. Kob, W. & Andersen, H. C. Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture—The Van Hove correlation-function. *Phys. Rev. E* **51,** 4626–4641 (1995).
23. Perera, D. N. & Harrowell, P. Relaxation dynamics and their spatial distribution in a two-dimensional glass-forming mixture. *J. Chem. Phys.* **111,** 5441–5454 (1999).
24. Murray, C. A. & Grier, D. G. Video microscopy of monodisperse colloidal systems. *Annu. Rev. Phys. Chem.* **47,** 421–462 (1996).
25. Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interf. Sci.* **179,** 298–310 (1996).
26. Kegel, W. K. & van Blaaderen, A. Direct observation of dynamical heterogeneities in colloidal hard-sphere suspensions. *Science* **287,** 290–293 (2000).
27. Weeks, E. R., Crocker, J. C., Levitt, A. C., Schofield, A. & Weitz, D. A. Three-dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science* **287,** 627–631 (2000).
28. Abate, A. R. & Durian, D. J. Approach to jamming in an air-fluidized granular bed. *Phys. Rev. E* **74,** 031308 (2006).
29. Busse, L. E. & Nagel, S. R. Temperature-dependence of the structure factor of As_2Se_3 glass up to the glass-transition. *Phys. Rev. Lett.* **47,** 1848–1851 (1981).
30. Lacevic, N., Starr, F. W., Schroder, T. B., Novikov, V. N. & Glotzer, S. C. Growing correlation length on cooling below the onset of caging in a simulated glass-forming liquid. *Phys. Rev. E* **66,** 030101 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Lubensky, D. Durian and K. Chen for discussions and a critical reading of the manuscript. We acknowledge the financial support of the Department of Energy and the National Science Foundation: DE-FG02-05ER46199 (A.J.L., N.X.), DE-FG02-03ER46088 (S.R.N., N.X.), the University of Chicago MRSEC DMR-0820054 (S.R.N., N.X.), DMR-080488 (A.G.Y.), and the PENN MRSEC DMR-0520020 (A.G.Y., A.J.L., Z.Z.). Z.Z. gratefully acknowledges partial support from Rhodia. Finally, we acknowledge the support of the Teraport computer cluster at the University of Chicago.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.Z. for experiments (zexin@sas.upenn.edu) or N.X. for simulations (ningxu@sas.upenn.edu).

Vol 459|14 May 2009|doi:10.1038/nature08003

LETTERS

White organic light-emitting diodes with fluorescent tube efficiency

Sebastian Reineke[1], Frank Lindner[1], Gregor Schwartz[1], Nico Seidler[1], Karsten Walzer[1], Björn Lüssem[1] & Karl Leo[1]

The development of white organic light-emitting diodes[1] (OLEDs) holds great promise for the production of highly efficient large-area light sources. High internal quantum efficiencies for the conversion of electrical energy to light have been realized[2–4]. Nevertheless, the overall device power efficiencies are still considerably below the 60–70 lumens per watt of fluorescent tubes, which is the current benchmark for novel light sources. Although some reports about highly power-efficient white OLEDs exist[5,6], details about structure and the measurement conditions of these structures have not been fully disclosed: the highest power efficiency reported in the scientific literature is 44 lm W^{-1} (ref. 7). Here we report an improved OLED structure which reaches fluorescent tube efficiency. By combining a carefully chosen emitter layer with high-refractive-index substrates[8,9], and using a periodic outcoupling structure, we achieve a device power efficiency of 90 lm W^{-1} at 1,000 candelas per square metre. This efficiency has the potential to be raised to 124 lm W^{-1} if the light outcoupling can be further improved. Besides approaching internal quantum efficiency values of one, we have also focused on reducing energetic and ohmic losses that occur during electron–photon conversion. We anticipate that our results will be a starting point for further research, leading to white OLEDs having efficiencies beyond 100 lm W^{-1}. This could make white-light OLEDs, with their soft area light and high colour-rendering qualities, the light sources of choice for the future.

To turn a white OLED into a power-efficient light source, three key parameters must be addressed: the internal electroluminescence quantum efficiency must be close to one (high internal quantum efficiency), a high fraction of the internally created photons must escape to the forward hemisphere (high outcoupling efficiency) and the energy loss during electron–photon conversion should be small (low operating voltage). The internal quantum efficiency and the outcoupling efficiency are combined in the external quantum efficiency (EQE).

The use of phosphors allows 100% internal quantum efficiency, because both the singlet and triplet states (generated at a ratio of 1:3 owing to their multiplicity) are directed to the emitting triplet state[10]. For power-efficient white OLEDs, an additional challenge is that high-energy phosphors demand host materials with even higher triplet energies to confine the excitation to the emitter[11]. Taking exciton binding energy and singlet–triplet splitting into account, the use of such host materials considerably increases the transport gap and therefore the operating voltage. For these reasons, blue fluorescent emitters are widely used to complete the residual phosphor-based emission spectrum[2,12,13]; this, however, either reduces the internal quantum efficiency or requires blue emitters with special properties[14]. Whenever OLEDs are built in a standard substrate emitting architecture, the outcoupling efficiency is approximately 20%. The remaining 80% of the photons are trapped in organic and substrate modes in equal amounts[15]. Hence, the greatest potential for a substantial increase in EQE and power efficiency is to enhance the light outcoupling.

Here we present an OLED structure that combines a novel concept for energy-efficient photon generation with improved outcoupling. The key feature of the OLED layer structure is the positioning of the blue phosphor within the emission layer and its combination with a carefully chosen host material: energetically, the triplet energy of the blue emitter material is in resonance with its host so that the blue phosphorescence is not accompanied by internal triplet energy relaxation before emission. The exciton formation region is at the interface of a double-emission-layer structure[16]. The blue host–guest system is surrounded by red and green sublayers of the emission layer to harvest unused excitons. For holes and electrons, the emission layer is nearly barrier-free until they reach the region of exciton formation, which keeps the operating voltage low. The outermost layers in contact with the electrodes are chemically p- and n-doped, which reduces ohmic losses to a negligible level[16].

A close-up of the emission layer (Fig. 1a) shows the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO)—the energy levels at which charge transport occurs—and the triplet energies of all materials. The latter essentially define the exciton distribution within the multilayer emission layer and, consequently, the emission spectrum and device efficiency. Holes and electrons are injected without facing any energy barrier into the emission layer from NPB to TCTA:Ir$(MDQ)_2$(acac) and from TPBi to TPBi:Ir$(ppy)_3$, respectively. (See Methods Summary for materials composition.) Here, holes are transported directly within the HOMO level of the emitter owing to its high concentration (10 wt%). Both carriers will accumulate at the double-emission-layer interface, forming excitons nearby. The different sublayers are separated by thin intrinsic interlayers of the corresponding host material to decouple the sublayers from unwanted energy transfer. Here, 2 nm is sufficient to suppress Förster-type transfer because the typical Förster radii for Ir complexes[17] are less than 2 nm. Excitons created in the blue region on host or dopant have various decay channels.

The transfer rate $k_{b\text{-}r}$ to the red emitter is strongly reduced with the introduction of the high-triplet-energy TCTA interlayer, restricting diffusive exciton migration[18]. Owing to their resonant triplet energies of 2.6 eV (see Fig. 2a), triplet excitons are free to move within the TPBi:FIrpic layer, resulting in a back-energy transfer rate k_{BT} accompanied by a delayed component in the decay of the emitting species[19]. This system cannot maintain the intrinsically high quantum yield of FIrpic, so the blue region is followed by an Ir$(ppy)_3$-doped region, retaining high efficiency by diffusively harvesting host excitons, represented by a rate of transfer from blue to green of $k_{b\text{-}g}$. The interlayer between blue and green ensures that solely diffusive energy exchange contributes to $k_{b\text{-}g}$, as Förster-type transfers are suppressed[17].

We now discuss the exciton dynamics in this emission layer. First, we present direct proof of the back-energy transfer k_{BT} in a complete

[1]Institut für Angewandte Photophysik, George-Bähr-Strasse 1, D-01062 Dresden, Germany.

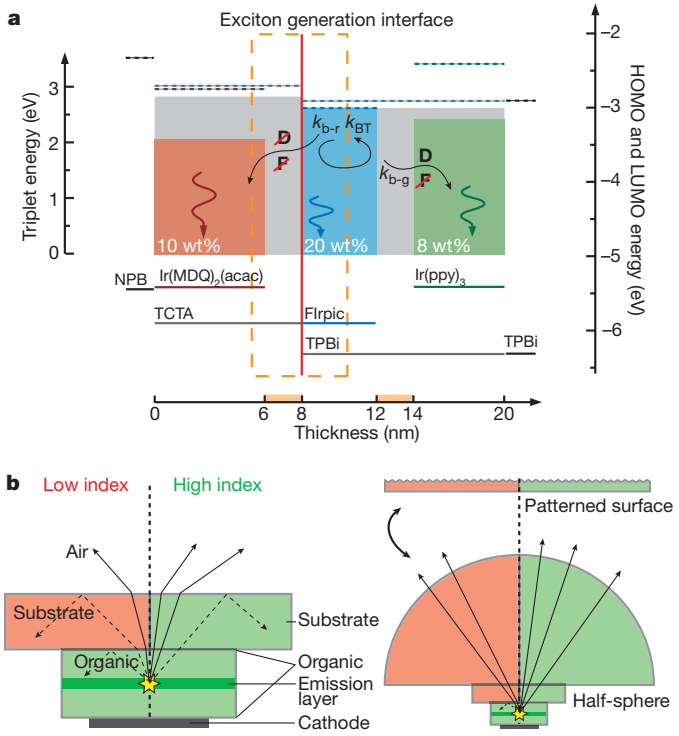


Figure 1 | Energy level diagram and light modes in an OLED. a, Lines correspond to HOMO (solid) and LUMO (dashed) energies; filled boxes refer to the triplet energies. The orange colour marks intrinsic regions of the emission layer. F and D represent Förster- and Dexter-type energy exchange channels, respectively. The orange dashed box depicts the main region of exciton generation. **b**, The left panel shows a cross-section of an OLED to illustrate the light propagation. Solid lines indicate modes escaping the device to the forward hemisphere; dashed lines represent trapped modes. The right panel shows how a large half-sphere and a patterned surface can be applied to increase light-outcoupling.

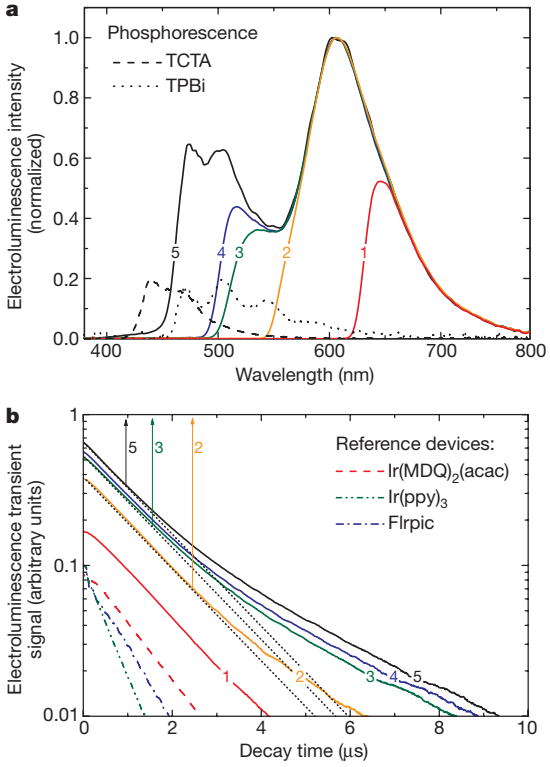


Figure 2 | Spectrum- and time-resolved electroluminescence transients. a, Electroluminescence spectra of device B obtained through colour filters. The spectra are numbered from solely red emission (1) to the complete emission (5). The phosphorescence spectra (at 77 K) of TCTA and TPBi are plotted. **b**, Electroluminescence decay curves of device B according to the spectrum in **a**. Arrows indicate the time at which a slower component sets in. From 1 to 5, this onset shifts to shorter times, nicely agreeing with a higher contribution of FIrpic emission. Additionally plotted are decays for blue, green and red reference devices.

device. This is followed by photoluminescence quantum yield measurements to confirm that excitons that cannot relax on FIrpic are captured diffusively by the green phosphor Ir(ppy)$_3$. Because k_{BT} is detected as a slow-relaxation component of the FIrpic emission, itself being one of several decay channels within the present white device structure (see Fig. 1a), we prepared an additional device B to increase the FIrpic emission, and hence k_{BT}, in the multicolour electroluminescence spectrum.

Figure 2 plots its spectrum- and time-resolved emission. In Fig. 2a, the emission is filtered using appropriate colour filters, starting with solely red emission (1) and subsequently increasing the transmission in the visible spectrum to the complete electroluminescence spectrum (5). The corresponding electroluminescence transients can be seen in Fig. 2b. First, a monoexponential decay with a time constant of 1.4 μs is observed for the red part of the spectrum (1). With increasing transmission, a second, slower component can be observed in the electroluminescence transient with a time constant of 3.0 μs. The spectral dependence, being directly linked to the FIrpic emission, indicates that this slow component can be exclusively attributed to k_{BT} from TPBi to FIrpic. The slow component is not seen for the blue reference device in Fig. 2b, because it comprises a TCTA:FIrpic emission layer, where excitons are confined to FIrpic[19].

The photoluminescence quantum yield η_{PL} is a very reliable measure of the suitability of emitter materials because it determines the ratio between the radiative decay channel (k_r) and the sum of radiative and non-radiative (k_{nr}) relaxation. In the present system, the rate of excitons relaxing without photon emission on host sites, k_H, needs to be included, making $\eta_{PL} = k_r/(k_r + k_{nr} + k_H)$. It is known that η_{PL} decreases for a phosphor-doped host–guest system whenever the excitation is not efficiently confined to the emitting species, and in most cases, this is accompanied by a back-energy transfer[11,19]. Measurements of η_{PL} are carried out to investigate the blue to green transfer[20] $k_{b\text{-}g}$.

FIrpic is doped at 1.7 wt% either into TCTA ($T_1 = 2.8$ eV, see Fig. 2a) or TPBi, yielding very different values for η_{PL} of 81% for TCTA and 14% for TPBi, indicating that TPBi:FIrpic alone cannot be used for an efficient OLED. TCTA, with a triplet energy about 0.2 eV higher, can efficiently confine excitons to FIrpic, resulting in a very high η_{PL} (here, $k_H = 0$). By knowing the triplet decay time, $\tau = 1/(k_r + k_{nr}) = 1.35$ μs, for the TCTA system[21], we can further deduce from the TPBi:FIrpic data that $k_H = 3.5 \times 10^6\ s^{-1}$. The latter is roughly six times larger than the radiative rate, $k_r = 6.0 \times 10^5\ s^{-1}$. It is the essence of this emission-layer design that these excitons are captured efficiently for green emission, that is, k_H feeds $k_{b\text{-}g}$ because the Ir(ppy)$_3$-doped region is within the triplet diffusion length of TPBi. The photoluminescence efficiency of TPBi:FIrpic increases to 32% when the FIrpic concentration is increased to 10 wt%. This indicates that the low η_{PL} of TPBi:FIrpic is not intrinsic, but instead depends on the probability of an exciton finding a dopant site for relaxation.

The use of high-refractive-index glass substrates can substantially increase the amount of light coupled from the organic layers into the glass substrate (up to 80%)[8,9]. In Fig. 1b, an OLED cross-section is shown to illustrate the light propagation originating in the emission layer. If we use a low-refractive-index substrate, light will face two interfaces with a step in the refractive index n. First, light will partly be reflected because of total internal reflection at the organic ($n_{org} = 1.7$–1.9)/glass substrate($n_{low} = 1.51$) interface, forming organic modes. Second, the light entering the glass substrate is facing the glass

substrate/air ($n_{air} = 1$) interface, where total internal reflection traps it to glass modes. Although organic modes remain inside the structure and therefore cannot contribute to the total light output of the device, glass modes can be coupled out by a modification of the substrate shape (see Fig. 1b). Increasing the refractive index of the glass substrate from $n_{low} = 1.51$ to $n_{high} = 1.78$ causes the index mismatch between organic materials and substrate to vanish, enhancing light coupling into the high-refractive-index glass, so that all photons guided to organic modes by total internal reflection at the organic/glass interface in the low-refractive-index case are entering the glass substrate.

Current density and luminance are plotted versus operating voltage for all devices in Fig. 3a, with the corresponding electroluminescence spectra displayed in Fig. 3b. For both substrate types, the OLEDs achieve a brightness of 1,000 cd m^{-2} slightly above 3 V; 10,000 cd m^{-2} are reached below 4 V. Devices LI and HI-1 exhibit an excellent colour-rendering index of 80, similar to the best values reported for white OLEDs[2,4,12,14]. The Commission Internationale d'Eclairage (CIE) coordinates of these devices are (0.44, 0.46) and (0.45, 0.47) for devices LI and HI-1, respectively. Because charges reach the emission layer almost without energetic barriers, the electroluminescence spectra of these devices do not depend on the brightness between 100 and 5,000 cd m^{-2}, which is a great improvement on many values from the literature (Supplementary Information)[2,14,22]. Figure 4 shows the power efficiencies of devices LI and HI-1, which differ only in the use of the substrate. Unless otherwise specified, all efficiency data throughout the text refer to a luminance in forward direction of 1,000 cd m^{-2}.

We obtained comparable power efficiencies of 30 and 33 lm W^{-1} without outcoupling enhancement, respectively, which corresponds to 13.1% EQE for device LI and 14.4% EQE for device HI-1. With the application of an index-matched glass half-sphere, device LI reaches 55 lm W^{-1} (24% EQE), which corresponds to an increase in EQE of a factor of 1.8. This relationship drastically changes for device HI-1. Here the EQE is increased by a factor of 2.4 to 34% EQE, corresponding to 81 lm W^{-1}. One promising approach to enhance light outcoupling even for large-area devices is the use of shaped substrates, which enables the coupling of light under high angles of incidence (to the substrate surface normal). We prepared a pattern of pyramids[9] (with period 0.5 mm) by cutting 90° grooves into a high index glass (Supplementary Information), similar to microlens arrays[23,24], to couple out more light. With the application of this patterned surface, device HI-1 achieves 26% EQE and 63 lm W^{-1}, already exceeding the values of device LI (with half-sphere). This result illustrates the great potential of high-refractive-index substrates.

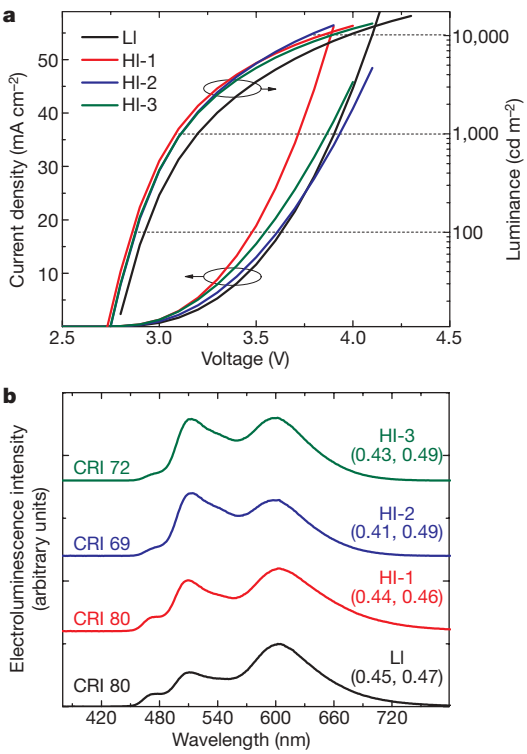


Figure 3 | Current density and luminance as a function of driving voltage and electroluminescence spectra of all devices. **a**, The data are obtained in the forward direction without outcoupling enhancement. Dashed lines indicate the brightness for 100, 1,000 and 10,000 cd m^{-2}. **b**, All data are obtained at 1,000 cd m^{-2}. Electroluminescence spectra, as displayed, are measured in direction normal to the glass substrate. In addition to the colour-rendering indices (CRI), the CIE coordinates are given. Both sets of data represent integrated values from the emission to the forward hemisphere.

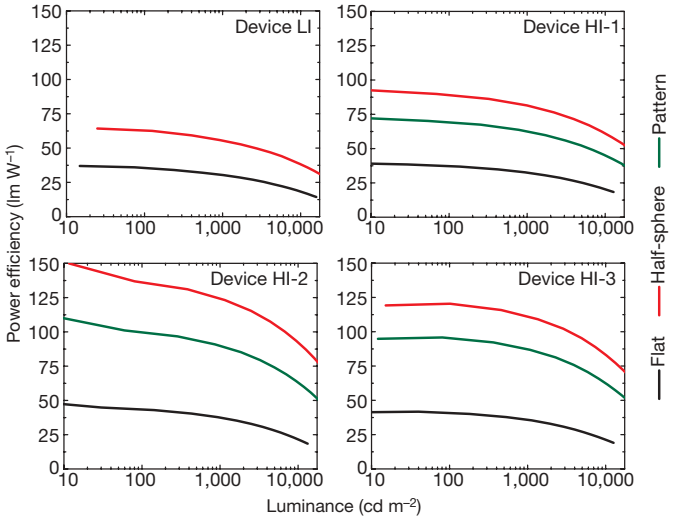


Figure 4 | Power efficiency of the white OLEDs. The power efficiencies of all devices are shown as a function of forward luminance. Black lines correspond to measurements of the planar structure (flat). The red lines are values obtained with additional index-matched half-spheres on top of the device, which indicate the maximum power efficiency because all light entering the glass substrate is coupled out to air. The green lines represent the measurements using a large-area light-outcoupling structure, that is, a periodically patterned index-matched substrate (made of Schott glass: N-LAF 21). For comparison, a typical fluorescent tube in a fixture reaches 60–70 lm W^{-1}.

The efficiency of organic LEDs can be increased further by placing the emission layer further away from the reflective cathode to avoid plasmonic losses to the metal[9,25]. Plasmonic losses, where the emitting dipoles couple to surface plasmons of the reflective metal, are the dominating loss channel when the emission takes place in the proximity of the metal. Their impact steadily decreases with greater distances between the emission layer and the cathode, and drops to a negligible level for distances greater than 200 nm (ref. 25). The light extraction to air is strongly influenced by the micro-cavity formed between glass and cathode, so we observe a periodical dependence of the emitted light as a function of distance between the emission layer and the cathode with maxima at distances corresponding to constructive interference of the emission wavelength. Additionally, if the emission layer is placed in the second antinode of the reflective metal cathode, OLEDs exhibit a more direct emission, which makes the light outcoupling of substrate modes easier (Supplementary Information)[25].

We prepared devices HI-2 and HI-3 with 205-nm-thick and 210-nm-thick electron transport layers, respectively, to best-fit the second outcoupling maximum. Their electroluminescence spectra are shown in Fig. 3b. Unlike devices LI and HI-1, we observed strong spectral changes. Here, emission from both the blue (FIrpic) and red ($Ir(MDQ)_2(acac)$) regions of the emitted spectrum was decreased,

negatively affecting both the colour rendering index (which decreases to ~70) and CIE coordinates (which shift into the yellow region (0.41–0.43, 0.49)). These changes can clearly be attributed to the different position of the second emission maximum for all three basic emitters, with a difference of roughly 60 nm for FIrpic and $Ir(MDQ)_2(acac)$ (Supplementary Information). This displacement from the Planck curve towards the yellow spectral range is not a large problem, and can be solved by using a deep blue phosphorescent emitter, which was not yet available to us.

The power efficiency of devices HI-2 and HI-3 can be seen in Fig. 4. Taking all substrate modes into account, these devices yield striking values of 124 and 111 lm W^{-1}, respectively. This corresponds to EQE values of 46% (HI-2) and 44% (HI-3), approaching efficiencies at which every second photon created is coupled into the forward hemisphere. Applying the pyramidal area structure to these devices, we obtain 90 lm W^{-1} (34% EQE) and 87 lm W^{-1} (34% EQE) for HI-2 and HI-3, respectively. These values are higher than the average power efficiency of fluorescent tubes in a fixture (60–70 lm W^{-1}). Furthermore, the novel emitter design is also characterized by an extremely small roll-off at high brightness (Supplementary Information): although it is common to state white OLED efficiency at 1,000 cd m^{-2}, higher brightness (2,000–5,000 cd m^{-2}) could significantly reduce the size and cost of OLED lighting. Such high brightness is usually challenging owing to the pronounced roll-off in efficiency, in particular for phosphorescent emitters[26], but at 5,000 cd m^{-2}, we obtain still very high power efficiencies of 74 lm W^{-1} (HI-2) and 73 lm W^{-1} (HI-3).

Our results show that white OLEDs with efficiencies approaching 100 lm W^{-1} even at high brightness are possible. For a broad application in general lighting, the lifetime issue of the blue emitters (Supplementary Information) has to be solved and the cost has to be significantly reduced, using low-cost electrode materials, thin-film encapsulation, roll-to-roll manufacturing and so on. With its potential to outperform fluorescent tubes, we think the future of white organic LEDs will be bright, not only because of their high illumination quality but also because their outstanding efficiencies will help to reduce our carbon footprint.

METHODS SUMMARY

All glass substrates were coated and structured with indium tin oxide (sheet resistance 25 Ω per square), and cleaned in an ultrasonic bath with acetone, ethanol and *iso*-propanol. All devices were fabricated by thermal evaporation in a single-chamber tool under high-vacuum conditions (base pressure ~10^{-8} mbar). Silver top contacts were thermally evaporated without breaking the vacuum. The devices were encapsulated with an additional glass and epoxy resin in a nitrogen atmosphere before evaluation. The device area is 6.7 mm^2.

The main materials used have acronyms as follows. MeO-TPD: *N*,*N*,*N*′,*N*′-tetrakis(4-methoxyphenyl)-benzidine. NPB: *N*,*N*′-di(naphthalen-1-yl)-*N*,*N*′)-diphenyl-benzidine. TPBi: 2,2′,2′′(1,3,5-benzenetriyl) tris-(1-phenyl-1H-benzimidazole). TCTA: 4,4′,4′′-tris(*N*-carbazolyl)-triphenylamine. Bphen: 4,7-diphenyl-1,10-phenanthroline. FIrpic: iridium-bis-(4,6,-difluorophenyl-pyridinato-N,C2)-picolinate. $[Ir(ppy)_3]$: *fac*-tris(2-phenylpyridine) iridium. $Ir(MDQ)_2$ (acac): iridium(III)bis(2-methyldibenzo[f,h]quinoxaline) (acetylacetonate).

The layer sequence for the white OLED on top of a low-index substrate (device LI) is as follows: 60 nm MeO-TPD doped with 4 mol.% NDP-2 as a hole-transport layer/10 nm NPB as the electron-blocker layer/emission layer/10 nm TPBi as a hole-blocking layer/40 nm Cs-doped Bphen as an electron-transport layer/100 nm Ag cathode. Alternatively, 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane can be used as freely available p-dopant (Supplementary Information). The emission layer (detailed composition is shown in Fig. 1) consists of a hole-transporting layer (TCTA) and an electron-transporting host material (TPBi) partially doped with the following phosphorescent emitters: FIrpic for blue, $[Ir(ppy)_3]$ for green and $Ir(MDQ)_2(acac)$ for orange.

Using a high-index glass substrate (devices HI-1, HI-2 and HI-3), the transport layers are adjusted to a hole-transport layer of 45 nm to optimize the light outcoupling. The thicknesses for the electron-transport layers are 40 nm (HI-1), 205 nm (HI-2) and 210 nm (HI-3), respectively. Unlike the standard emission layer, the thickness of the TPBi:FIrpic sublayer is increased from 4 to 8 nm in device B to enhance the FIrpic emission. Device efficiencies were measured in a calibrated integrating sphere. HOMO values are obtained from ultraviolet photoelectron spectroscopy; LUMO values are estimated from the optical gap of the material[16,27,28].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 December 2008; accepted 20 March 2009.

1. Kido, J., Kimura, M. & Nagai, K. Multilayer white light-emitting organic electroluminescent device. *Science* **267,** 1332–1334 (1995).
2. Sun, Y. *et al.* Management of singlet and triplet excitons for efficient white organic light-emitting devices. *Nature* **440,** 908–912 (2006).
3. Williams, E. L., Haavisto, K., Li, J. & Jabbour, G. E. Excimer-based white phosphorescent organic light emitting diodes with nearly 100% internal quantum efficiency. *Adv. Mater.* **19,** 197–202 (2007).
4. Sun, Y. & Forrest, S. R. High-efficiency white organic light emitting devices with three separate phosphorescent emission layers. *Appl. Phys. Lett.* **91,** 263503 (2007).
5. Nakayama, T., Hiyama, K., Furukawa, K. & Ohtani, H. Development of phosphorescent white OLED with extremely high power efficiency and long lifetime. *SID07 Dig.* 1018–1021 (2006).
6. D'Andrade, B. W. *et al.* Realizing white phosphorescent 100 lm/W OLED efficacy. *Proc. SPIE* **7051,** 70510Q (2008).
7. Su, S.-J., Gonmori, E., Sasabe, H. & Kido, J. Highly efficient organic blue-and white-light-emitting devices having a carrier- and exciton-confining structure for reduced efficiency roll-off. *Adv. Mater.* **20,** 4189–4195 (2008).
8. Nakamura, T., Tsutsumi, N., Juni, N. & Fujii, H. Thin-film waveguiding mode light extraction in organic electroluminescent device using high refractive index substrate. *Appl. Phys. Lett.* **97,** 054505 (2005).
9. Gärtner, G. & Greiner, H. Light extraction from OLEDs with (high) index matched glass substrates. *Proc. SPIE* **6999,** 69992T (2008).
10. Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395,** 151–154 (1998).
11. Goushi, K., Kwong, R., Brown, J. J., Sasabe, H. & Adachi, C. Triplet exciton confinement and unconfinement by adjacent hole-transport layers. *J. Appl. Phys.* **95,** 7798–7802 (2004).
12. Schwartz, G., Fehse, K., Pfeiffer, M., Walzer, K. & Leo, K. Highly efficient white organic light emitting diodes comprising an interlayer to separate fluorescent and phosphorescent regions. *Appl. Phys. Lett.* **89,** 083509 (2006).
13. Qin, D. & Tao, Y. White organic light-emitting diode comprising of blue fluorescence and red phosphorescence. *Appl. Phys. Lett.* **86,** 113507 (2005).
14. Schwartz, G., Pfeiffer, M., Reineke, S., Walzer, K. & Leo, K. Harvesting triplet excitons from fluorescent blue emitters in white organic light-emitting diodes. *Adv. Mater.* **19,** 3672–3676 (2007).
15. Greenham, N. C., Friend, R. H. & Bradley, D. D. C. Angular-dependence of the emission from a conjugated polymer light-emitting diode - implications for efficiency calculations. *Adv. Mater.* **6,** 491–494 (1994).
16. He, G. *et al.* High-efficiency and low-voltage p-i-n electrophosphorescent organic light-emitting diodes with double-emission layers. *Appl. Phys. Lett.* **85,** 3911–3913 (2004).
17. Kawamura, Y., Brooks, J., Brown, J. J., Sasabe, H. & Adachi, C. Intermolecular interaction and a concentration-quenching mechanism of phosphorescent Ir(III) complexes in a solid film. *Phys. Rev. Lett.* **96,** 017404 (2006).
18. Reineke, S., Schwartz, G., Walzer, K. & Leo, K. Reduced efficiency roll-off in phosphorescent organic light emitting diodes by suppression of triplet-triplet annihilation. *Appl. Phys. Lett.* **91,** 123508 (2007).
19. Kawamura, Y. *et al.* 100% phosphorescence quantum efficiency of Ir(III) complexes in organic semiconductor films. *Appl. Phys. Lett.* **86,** 071104 (2005).
20. Greenham, N. C. *et al.* Measurement of absolute photoluminescence quantum efficiencies in conjugated polymers. *Chem. Phys. Lett.* **241,** 89–96 (1995).
21. Huang, Q., Reineke, S., Walzer, K., Pfeiffer, M. & Leo, K. Quantum efficiency enhancement in top-emitting organic light-emitting diodes as a result of enhanced intrinsic quantum yield. *Appl. Phys. Lett.* **89,** 263512 (2006).
22. D'Andrade, B. W., Holmes, R.-J. & Forrest, S. R. Efficient organic electrophosphorescent white-light-emitting device with a triple doped emissive layer. *Adv. Mater.* **16,** 624–628 (2004).
23. Madigan, C. F., Lu, M.-H. & Sturm, J. C. Improvement of output coupling efficiency of organic light-emitting diodes by backside substrate modification. *Appl. Phys. Lett.* **76,** 081114 (2006).
24. Möller, S. & Forrest, S. R. Improved light out-coupling in organic light emitting diodes employing ordered microlens arrays. *J. Appl. Phys.* **91,** 3324–3327 (2002).
25. Lin, C.-H., Cho, T.-Y., Chang, C.-H. & Wu, C.-C. Enhancing light outcoupling of organic light-emitting devices by locating emitters around the second antinode of the reflective metal electrode. *Appl. Phys. Lett.* **88,** 081114 (2006).
26. Baldo, M. A., Adachi, C. & Forrest, S. R. Transient analysis of organic electrophosphorescence: II. Transient analysis of triplet-triplet annihilation. *Phys. Rev. B* **62,** 10967 (2000).
27. Meerheim, R. *et al.* Influence of charge balance and exciton distribution on efficiency and lifetime of phosphorescent organic light-emitting devices. *Appl. Phys. Lett.* **104,** 014510 (2008).

28. Adamovich, V. *et al.* New charge-carrier blocking materials for high efficiency OLEDs. *Org. Electron.* **4,** 77–87 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the European Commission within the sixth framework IST programme under contract IST-2002-004607 (project OLLA), for funding. We received further support via the Leibniz Prize of the Deutsche Forschungsgemeinschaft. We also thank Novaled AG, Dresden, for providing the hole-transport layer dopant NDP-2 as well as J. Förster and T. Günther for technical assistance throughout sample preparation.

Author Contributions S.R. designed the emission concept, performed the transient electroluminescence measurements, wrote the manuscript, analysed most of the data and, together with F.L., optimized and characterized the devices and designed the outcoupling structure. G.S. was involved in the development of the second maximum devices. N.S. performed the photoluminescence quantum yield measurements. K.W. and B.L. coordinated the high efficiency white OLED project. K.L. motivated this work and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.L. (leo@iapp.de).

METHODS

Reference devices. The electroluminescent decay curves of Fig. 2b correspond to the following reference samples prepared on top of standard glass coated with indium tin oxide. FIrpic consists of: 60 nm MeO-TPD:NDP-2/10 nm NPB/20 nm TCTA:FIrpic 20 wt% (ref. 21)/10 nm TPBi/50 nm Bphen:Cs/100 nm Al. $Ir(ppy)_3$ consists of: 60 nm MeO-TPD:NDP-2/10 nm NPB/20 nm TCTA:$Ir(ppy)_3$ 8 wt% (ref. 18)/10 nm TPBi/50 nm Bphen:Cs/100 nm Al. $Ir(MDQ)_2(acac)$ consists of: 60 nm MeO-TPD:NDP-2/10 nm NPB/20 nm NPB:$Ir(MDQ)_2(acac)$ 10 wt% (ref. 27)/10 nm TPBi/50 nm Bphen:Cs/100 nm Al.

Device evaluation. Electroluminescence spectra were recorded with a calibrated spectrometer CAS 140 CT (Instrument Systems Optische Messtechnik). Only the electroluminescence spectra as shown in Fig. 2 were recorded with a USB2000 minispectrometer (OceanOptics). All efficiency measurements were carried out in an integrating sphere (Instrument Systems Optische Messtechnik) attached to the calibrated spectrometer CAS 140 CT and a source-measure unit 2400 (Keithley Instruments). The relative efficiencies as a function of luminance were measured with a fast, calibrated photodetector in the forward direction, which were then rescaled to the values obtained with the integrating sphere. This is valid because the electroluminescence spectra do not change significantly in the displayed range of brightness. All efficiencies are given, if not stated otherwise, at a luminance of 1,000 cd m^{-2} and measured in the forward direction, that is, at a normal angle of incidence for the complete device configuration, eventually including outcoupling structures. The glass half-spheres have diameters of 18 and 15 mm for low and high refractive index, respectively. Index-matching oils of $n = 1.5$ and $n = 1.78$ were obtained from Olympus Corporation and Cargille Laboratories, respectively. Substrate edges were covered to exclude edge emission contributing to the measurement. Photoluminescence quantum yield measurements were carried out in an integrating sphere (Labsphere) using a 325 nm HeCd laser (Kimmon Electric Company) as excitation source and the USB2000 minispectrometer as detector. The set-up was calibrated using a ultraviolet/visible light source, itself calibrated with the CAS 140 CT spectrometer.

Spectroscopy. The phosphorescence spectra of TCTA and TPBi in Fig. 2a were measured at 77 K using a gated phosphorescence set-up with a 337 nm pulsed laser (MSG-SD from Lasertechnik Berlin) as excitation source. Here, the delay generator (DG 535, Stanford Research Systems), triggered with the laser pulse, gave the delay for the detection (LS 50 B spectrometer, Perkin Elmer), to separate fast and slow phosphorescence. The time and spectrally resolved measurements were carried out under electroluminescence operation. The general set-up is realized as shown previously[29]. Using different colour filters, the transmission of the white OLED spectrum was changed. The transmitted intensity was then linked to the fast photodiode with a glass fibre to detect the time decay.

29. Reineke, S. *et al.* Measuring carrier mobility in conventional multilayer organic light emitting devices by delayed exciton generation. *Phys. Status Solidi B* **245**, 804–809 (2008).

LETTERS

Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions

Matthew W. Powner[1], Béatrice Gerland[1] & John D. Sutherland[1]

At some stage in the origin of life, an informational polymer must have arisen by purely chemical means. According to one version of the 'RNA world' hypothesis[1–3] this polymer was RNA, but attempts to provide experimental support for this have failed[4,5]. In particular, although there has been some success demonstrating that 'activated' ribonucleotides can polymerize to form RNA[6,7], it is far from obvious how such ribonucleotides could have formed from their constituent parts (ribose and nucleobases). Ribose is difficult to form selectively[8,9], and the addition of nucleobases to ribose is inefficient in the case of purines[10] and does not occur at all in the case of the canonical pyrimidines[11]. Here we show that activated pyrimidine ribonucleotides can be formed in a short sequence that bypasses free ribose and the nucleobases, and instead proceeds through arabinose amino-oxazoline and anhydronucleoside intermediates. The starting materials for the synthesis—cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate—are plausible prebiotic feedstock molecules[12–15], and the conditions of the synthesis are consistent with potential early-Earth geochemical models. Although inorganic phosphate is only incorporated into the nucleotides at a late stage of the sequence, its presence from the start is essential as it controls three reactions in the earlier stages by acting as a general acid/base catalyst, a nucleophilic catalyst, a pH buffer and a chemical buffer. For prebiotic reaction sequences, our results highlight the importance of working with mixed chemical systems in which reactants for a particular reaction step can also control other steps.

Because they comprise phosphate, ribose and nucleobases, it is tempting to assume that ribonucleotides must have prebiotically assembled from such building blocks. Thus, for example, it has previously been supposed that the activated ribonucleotide β-ribocytidine-2′,3′-cyclic phosphate **1** must have been produced by phosphorylation of the ribonucleoside **2**, with the latter deriving from the conjoining of the free pyrimidine nucleobase cytosine **3** and the furanose form of ribose **4** (Fig. 1, blue arrows). This mode of assembly is seemingly supported by the facts that cytosine **3** can be synthesized by condensation of cyanoacetaldehyde **5** and urea **6**[16] (the hydration products of cyanoacetylene **7**[17], and cyanamide **8**[18], respectively) and pentoses including ribose can be produced by aldol reaction of glyceraldehyde **9** and glycolaldehyde **10**[8,9]. The insuperable problem with this approach, however, is that one of the presumed steps, the condensation of ribose **4** and cytosine **3**, does not work[11]. The reasons for this are both kinetic (the N1 lone pair of **3** is unavailable owing to delocalization) and, in water, thermodynamic (the equilibrium constant is such that hydrolysis of **2** to **3** and **4** is favoured over condensation). The same is true for ribosylation of uracil, which has also not been demonstrated.

We have considered a large number of alternative ribonucleotide assembly modes, including those that extend back to the same small-molecule precursors as the traditionally assumed route described above[19]. By systematic experimental investigation of these options, we have discovered a short, highly efficient route to activated pyrimidine ribonucleotides from these same precursors that proceeds by way of alternative intermediates (Fig. 1, green arrows). By contrast with previously investigated routes to ribonucleotides, ours bypasses ribose and the free pyrimidine nucleobases. Mixed nitrogenous–oxygenous chemistry first results in the reaction of cyanamide **8** and glycolaldehyde **10**, giving 2-amino-oxazole **11**, and this heterocycle then adds to glyceraldehyde **9** to give the pentose amino-oxazolines including the arabinose derivative **12**. Reaction of **12** with cyanoacetylene **7** then gives the anhydroarabinonucleoside **13**, which subsequently undergoes phosphorylation with rearrangement to furnish β-ribocytidine-2′,3′-cyclic phosphate **1**. In a subsequent photochemical step, **1** is partly converted to the corresponding uracil derivative, and synthetic co-products are largely destroyed.

We had previously shown that in unbuffered aqueous solution, 2-amino-oxazole **11** adds to glyceraldehyde **9** to give the pentose amino-oxazolines including **12** in excellent overall yield[20]. Our starting point in the present work was therefore to find a prebiotically

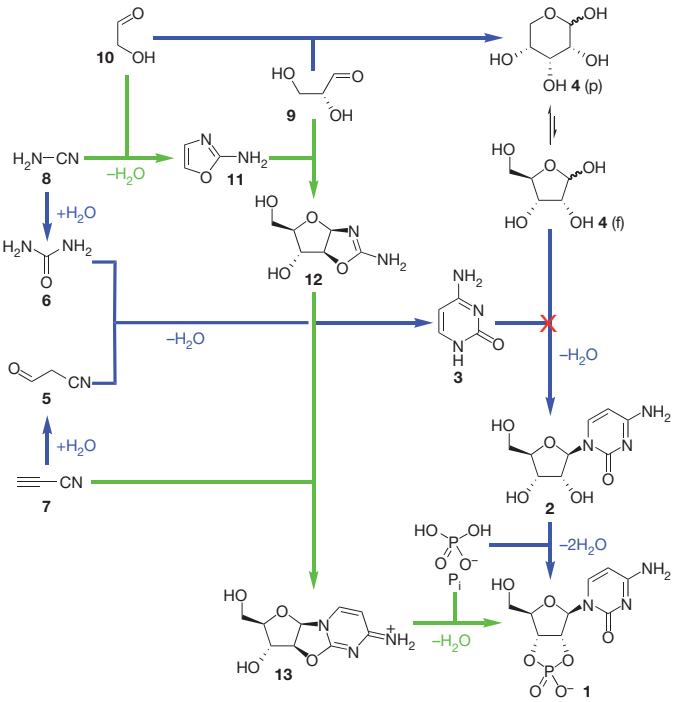


Figure 1 | Pyrimidine ribonucleotide assembly options. Previously assumed synthesis of β-ribocytidine-2′,3′-cyclic phosphate **1** (blue; note the failure of the step in which cytosine **3** and ribose **4** are proposed to condense together) and the successful new synthesis described here (green). p, pyranose; f, furanose.

[1]School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, UK.

plausible synthesis of **11**. Constitutionally, **11** is the condensation product of **8** and **10**, and although there exists, in the conventional chemical literature, a procedure to bring about this condensation, it requires strongly alkaline conditions[21]. Because we wanted to generate **11**, and then allow it to react with **9**, which is unstable to alkali, under the same conditions, neutral-pH reaction conditions had to be found.

We initially investigated the reaction with **8** and **10** in a 1:1 ratio starting at neutral pH in unbuffered aqueous solution. Only a small amount of **11** was produced under these conditions, and ^{1}H NMR spectra were indicative of the formation of a variety of carbonyl addition adducts and other intermediates, for example **14**–**18** (Fig. 2a, b). The carbonyl addition adducts **14** were presumably formed reversibly, and so did not represent material irretrievably committed to other products, but rather intermediates stalled en route to **11**. At low concentrations of hydroxide, it appeared that two additional types of reaction needed to make **11** were very sluggish: intra-adduct attack of the glycolaldehyde-derived hydroxyl group on the cyanamide-derived nitrile carbon (for example **14** ($n = 0$) → **15**), and C–H deprotonation leading to aromatization (**17** → **11**).

Denied the opportunity of using hydroxide as a specific base catalyst to accelerate these slow steps, we sought a general base catalyst that could provide the same acceleration, but at neutral pH. Inorganic phosphate seemed to be ideal in this regard because its second pK_a value is close to neutrality. Furthermore, as phosphate is ultimately needed in some form to make activated nucleotides, we decided to include it from the start of the assembly sequence. We repeated the earlier reaction of cyanamide **8** and glycolaldehyde **10**, but in the presence of 1 M phosphate buffer at pH 7.0. ^{1}H NMR analysis revealed that 2-amino-oxazole **11** was produced in >80% yield (75% isolated yield) (Fig. 2c). With an excess of **8** over **10**, the synthesis of **11** still takes place in the presence of phosphate, but is followed by slower phosphate addition to residual **8** giving the intermediate adduct **19**, which partitions to urea **6** and cyanoguanidine **20** (Fig. 2d).

We then investigated whether the subsequent reaction of **11** with glyceraldehyde **9** would be tolerant to the residual presence of phosphate. In the absence of phosphate, the ribose and arabinose amino-oxazolines **21** and **12** are the major products, and the xylose derivative **22** is a minor product (Fig. 3a)[20]. The lyxose amino-oxazoline **23** is formed in intermediate amounts as an equilibrating mixture of pyranose and furanose isomers. All of the pentose amino-oxazolines have the potential to be converted reversibly into one or other of the 5-substituted 2-amino-oxazoles **24** and **25** by phosphate catalysis (by chemistry similar to that underlying the conversion of **16** to **11**), but to differing extents depending on their stability. After one day in the presence of phosphate, all of the amino-oxazolines showed some conversion to the corresponding 5-substituted 2-amino-oxazole (**24** or **25**), but the lyxose amino-oxazoline **23** proved the least stable and underwent the greatest conversion (Fig. 3b). We then took a crude sample of **11** that had just been prepared from cyanamide **8** and glycolaldehyde **10** in the presence of phosphate, and added glyceraldehyde **9** to it. After overnight incubation, ^{1}H NMR analysis revealed that although all four amino-oxazolines were still formed, the lyxose derivative **23** was selectively depleted and was now a minor product along with the xylose derivative **22** (Fig. 3c). With two of its stereoisomeric relatives now minor products, the path from the arabinose amino-oxazoline **12** to ribonucleotides looked clearer. Selective crystallization of ribose amino-oxazoline **21** offers a further means of enriching **12** such that it becomes the major product in solution[20,22].

We then proceeded to the second stage of pyrimidine nucleobase assembly. Although our focus was on the chemistry of the key arabinose amino-oxazoline **12**, the corresponding chemistry of the ribose amino-oxazoline **21** was also studied (Supplementary Information). It had earlier been shown that in unbuffered aqueous solution, **12** reacts with an excess of cyanoacetylene **7** giving β-arabinocytidine **26**, (Fig. 4a)[23]. The yield of **26** was relatively low, however, and we used ^{1}H NMR analysis to determine why. It transpires that the pH rises during the course of the reaction, resulting in hydrolysis of anhydronucleoside

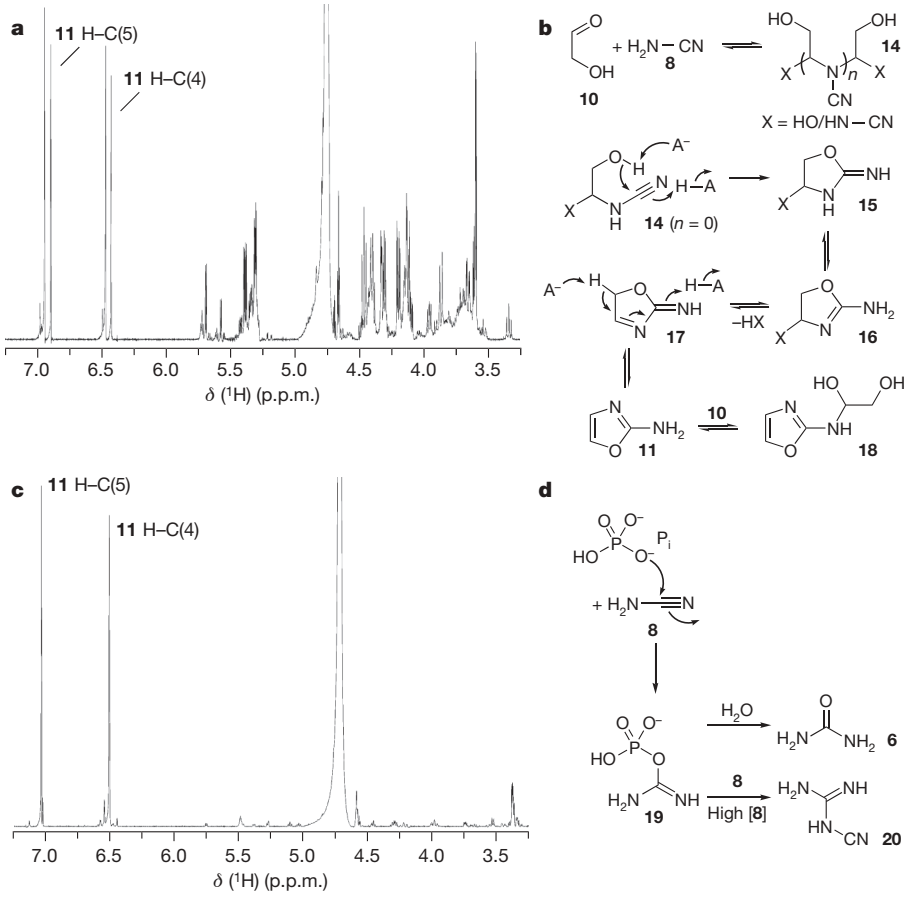


Figure 2 | Development of the synthesis of 2-amino-oxazole 11. a, ^{1}H NMR spectrum of the products of reaction of cyanamide **8** and glycolaldehyde **10** in the absence of phosphate. δ, chemical shift. **b**, Mechanism for the condensation of **8** and **10**. Curved arrows depict base-catalysed steps thought to be rate limiting in the absence of phosphate. **c**, ^{1}H NMR spectrum of the products of reaction of **8** and **10** in the presence of phosphate. **d**, Slower side reaction between **8** and phosphate. P_i, inorganic phosphate; H–A, general acid; A^-, general base.

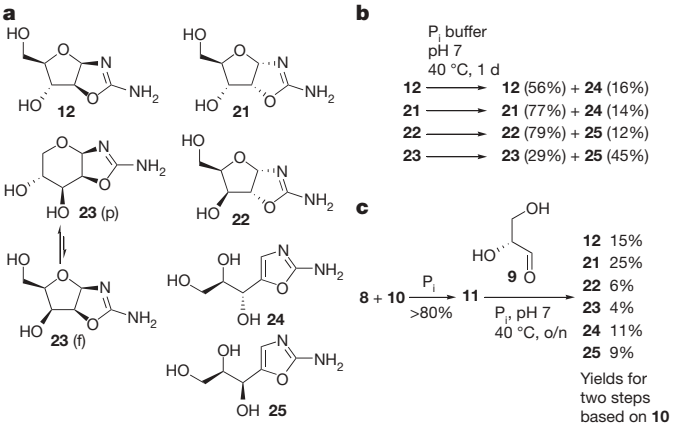


Figure 3 | Pentose amino-oxazoline stability, and assembly chemistry. **a**, Structures of the arabinose (**12**), ribose (**21**), xylose (**22**) and lyxose (**23**) amino-oxazolines and their elimination products **24** and **25**. **b**, Relative stabilities of the amino-oxazolines in the presence of phosphate. **c**, Formation of amino-oxazolines by addition of glyceraldehyde **9** to a solution of 2-amino-oxazole **11**, with the latter freshly formed *in situ* from cyanamide **8** and glycolaldehyde **10**. P_i, inorganic phosphate; o/n, overnight.

intermediates and causing hydroxyl groups to undergo reaction with cyanoacetylene **7** (Supplementary Information). To prevent the rise in pH during the reaction, inorganic phosphate was added as a buffer. When the buffering pH was 6.5, the reactions were extremely clean, with little evidence for anhydronucleoside hydrolysis. Furthermore, excess cyanoacetylene **7** that did not evaporate underwent reaction with phosphate at this pH, giving cyanovinyl phosphate **27** instead of cyanovinylating hydroxyl groups. Using phosphate as a dual-function pH and chemical buffer in this way, the arabinose anhydronucleoside **13** could be produced in extremely high yield from **12**.

Our finding that the reaction of the amino-oxazoline **12** with cyanoacetylene **7** could be controlled, by the pH and chemical buffering action of phosphate, to produce the arabinose anhydronucleoside **13** in excellent yield opened up the possibility of a combined phosphorylation–rearrangement[24,25] reaction to convert **13** to the activated ribonucleotide **1**. Furthermore, the formation of cyanovinyl phosphate **27**, as a co-product in the nucleobase assembly process, extended the range of potential phosphorylating agents for such a process because, in aqueous solution, **27** undergoes reaction with inorganic phosphate to give pyrophosphate[17]. Accordingly, we investigated the phosphorylation of the anhydronucleoside **13** using both inorganic phosphate and pyrophosphate.

Prebiotic phosphorylation of nucleosides has been demonstrated by heating either in the dry state with urea[26] or in formamide solution[27]. We were particularly attracted by the possibility of using urea **6** in the phosphorylation of **13** because it is a co-product of the chemical system in which 2-amino-oxazole **11** is produced from glycolaldehyde **10** and cyanamide **8**, if the latter is initially present in excess (Fig. 2d). After preliminary experiments, and through consideration of the phosphorylation mechanism (Supplementary Information), we found that when **13** was heated with 0.5 equiv. of pyrophosphate in urea containing ammonium salts, **1** was formed as the major product in addition to **28** and **29** (the 5′-phosphate derivatives of **13** and **1**, respectively) and small amounts of the hydrolysis product β-arabinocytidine **26** and its 5′-phosphate derivative **30** (Fig. 4b, procedure A; Supplementary Information). Alternatively, **1** was formed in very good yield—along with **29**, the hydrolysis products **26** and **30**, and the nucleobases cytosine **3** and diaminopyrimidine **31**—by heating **13** with inorganic phosphate and urea in formamide solution (Fig. 4b, procedure B; Supplementary Information).

The conversion to **1** in both cases is thought to involve phosphorylation of the 3′-hydroxyl group of **13** to give the 3′-phosphate **32**, which can undergo rearrangement, through intramolecular nucleophilic substitution (Fig. 4c)—a reaction not possible in the *ribo*-analogue because of the *cis* relationship of the 2′- and 3′-oxygens. The efficiency of the conversion of **13** to **1** is thought to be due to a high selectivity for 3′-phosphorylation over 5′-phosphorylation. Such selectivity is particularly noteworthy because of the increased steric hindrance normally associated with a secondary alcohol in comparison with a primary alcohol. To investigate this, we determined the X-ray crystal structure of **13**, and found that its sugar moiety has the C(4′)-*endo* pucker with a C(4′)–C(5′) +*sc* conformation (Fig. 4d). This conformation has the effect of making the 5′-hydroxyl group of **13** abnormally hindered for a nucleoside derivative relative to the 3′-hydroxyl group. Assuming that the solid-state conformation of **13** is also the predominant conformation in urea and in formamide solution, the relative ease with which the 3′- and

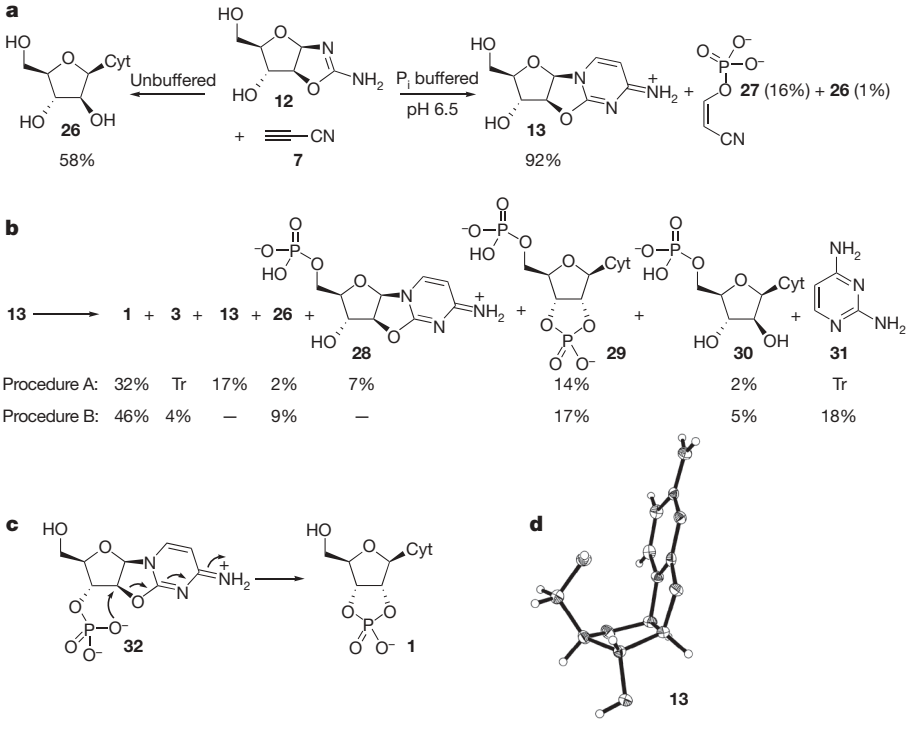


Figure 4 | Formation and phosphorylation of the arabinose anhydronucleoside 13. **a**, Major products isolated from the reaction of arabinose amino-oxazoline **12** and cyanoacetylene **7** in the presence and absence of phosphate. **b**, Products of phosphorylation of **13** using pyrophosphate and urea in the dry state (procedure A) or inorganic phosphate and urea in formamide solution (procedure B) (see main text and Supplementary Information). **c**, Rearrangement of **32**, the 3′-phosphate of **13**, to **1** by intramolecular nucleophilic substitution. **d**, X-ray crystal structure of **13**. Cyt, N1-linked cytosine; Tr, trace.

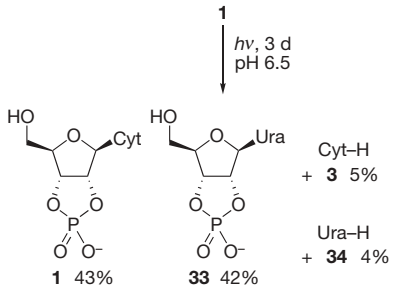


Figure 5 | Photochemistry of β-ribocytidine-2′,3′-cyclic phosphate 1. Under conditions of irradiation that destroy most other pyrimidine nucleosides and nucleotides (Supplementary Information), 1 undergoes partial hydrolysis and slight nucleobase loss. Ura, N1-linked uracil; Cyt–H, cytosine; Ura–H, uracil.

5′-hydroxyl groups are phosphorylated can be understood on the basis of these conformationally controlled steric effects.

If the products of the phosphorylation reaction in urea were subsequently to dissolve in aqueous medium at neutral pH and incubate for any significant length of time, then any residual anhydronucleoside/anhydronucleotide would undergo hydrolysis. Assuming such a rehydration, after phosphorylation in urea or urea–formamide mixtures, the major nucleosides/nucleotides that would accompany **1** would be **26**, **29** and **30** (in addition to any products ultimately deriving from the ribose amino-oxazoline **21** that was a by-product in the synthesis of arabinose amino-oxazoline **12**; see Supplementary Information). It is apparent that although **1** would be one of the major products, these co-products might interfere with any subsequent incorporation of **1** into RNA. Accordingly, we sought a means of selectively destroying these co-products. Furthermore, we also hoped to find a way of converting **1** partly to the corresponding activated uracil nucleotide, β-ribouridine-2′,3′-cyclic phosphate **33**. It transpires that irradiation achieves both of these goals.

Limited irradiation of aqueous solutions of cytosine nucleosides with ultraviolet light having an emission maximum at 254 nm results in the reversible formation of photohydrates and partial hydrolysis to the corresponding uracil nucleosides[28]. Prolonged irradiation causes additional chemistry to take place[29], and results in the destruction of most pyrimidine nucleosides and nucleotides (for example **26**, **30** and the major nucleoside/nucleotide products deriving from ribose amino-oxazoline **21**; see Supplementary Information). By contrast, however, we found that prolonged irradiation of β-ribocytidine-2′,3′-cyclic phosphate **1** causes significant hydrolysis to β-ribouridine-2′,3′-cyclic phosphate **33**, with very little destructive photochemistry other than slight nucleobase loss; cytosine **3** and uracil **34** were both detected (Fig. 5). This finding suggests that there must be some protective mechanism functioning with **1** and **33** that does not operate with other pyrimidine nucleosides and nucleotides. Whatever the mechanism (Supplementary Information), the protection against the destructive effects of irradiation provides a means whereby **1** and **33**, the two activated pyrimidine ribonucleotides needed for RNA synthesis, can be enriched relative to other end products of the assembly process we have discovered.

Our findings suggest that the prebiotic synthesis of activated pyrimidine nucleotides should be viewed as predisposed[30]. This predisposition would have allowed the synthesis to operate on the early Earth under geochemical conditions suitable for the assembly sequence. Although the issue of temporally separated supplies of glycolaldehyde and glyceraldehyde remains a problem, a number of situations could have arisen that would result in the conditions of heating and progressive dehydration followed by cooling, rehydration and ultraviolet irradiation. Comparative assessment of these models is beyond the scope of this work, but it is hoped that the chemistry described here will contribute to such an assessment.

Received 11 December 2008; accepted 24 March 2009.

1. Woese, C. *The Genetic Code* 179–195 (Harper & Row, 1967).
2. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38,** 367–379 (1968).
3. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–393 (1968).
4. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418,** 214–221 (2002).
5. Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 23–56 (Cold Spring Harbor Laboratory Press, 2006).
6. Ferris, J. P., Hill, A. R. Jr, Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381,** 59–61 (1996).
7. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol.* **2,** 303–316 (1973).
8. Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3,** 1850–1855 (2005).
9. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303,** 196 (2004).
10. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67,** 25–33 (1972).
11. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123 (2004).
12. Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Phil. Trans. R. Soc. B* **361,** 1681–1687 (2006).
13. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Cyanoacetylene in prebiotic synthesis. *Science* **154,** 784–785 (1966).
14. Pasek, M. A. & Lauretta, D. S. Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* **5,** 515–535 (2005).
15. Bryant, D. E. & Kee, T. P. Direct evidence for the availability of reactive, water soluble phosphorus on the early Earth. H-phosphinic acid from the Nantan meteorite. *Chem. Commun.* 2344–2346 (2006).
16. Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **373,** 772–774 (1995).
17. Ferris, J. P., Goldstein, G. & Beaulieu, D. J. Chemical evolution, IV. An evaluation of cyanovinyl phosphate as a prebiotic phosphorylating agent. *J. Am. Chem. Soc.* **92,** 6598–6603 (1970).
18. Lohrmann, R. & Orgel, L. E. Prebiotic synthesis: phosphorylation in aqueous solution. *Science* **161,** 64–66 (1968).
19. Anastasi, C. *et al.* RNA: prebiotic product, or biotic invention? *Chem. Biodivers.* **4,** 721–739 (2007).
20. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chem. Int. Edn Engl.* **45,** 6176–6179 (2006).
21. Cockerill, A. F. *et al.* An improved synthesis of 2-amino-1,3-oxazoles under basic catalysis. *Synthesis* 591–593 (1976).
22. Springsteen, G. & Joyce, G. F. Selective derivatization and sequestration of ribose from a prebiotic mix. *J. Am. Chem. Soc.* **126,** 9578–9583 (2004).
23. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47,** 531–543 (1970).
24. Tapiero, C. M. & Nagyvary, J. Prebiotic formation of cytidine nucleotides. *Nature* **231,** 42–43 (1971).
25. Shannahoff, D. H. & Sanchez, R. A. 2,2′-Anhydronucleosides. Novel syntheses and reactions. *J. Org. Chem.* **38,** 593–598 (1973).
26. Lohrmann, R., &. Orgel, L. E. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* **171,** 490–494 (1971).
27. Schoffstall, A. M. Prebiotic phosphorylation of nucleosides in formamide. *Orig. Life* **7,** 399–412 (1976).
28. Miller, N. & Cerutti, P. Structure of the photohydration products of cytidine and uridine. *Proc. Natl Acad. Sci. USA* **59,** 34–38 (1968).
29. Powner, M. W. *et al.* On the prebiotic synthesis of ribonucleotides: photoanomerisation of cytosine nucleosides and nucleotides revisited. *ChemBioChem* **8,** 1170–1179 (2007).
30. Eschenmoser, A. & Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem. Soc. Rev.* **21,** 1–16 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work has been funded by the UK Engineering and Physical Sciences Research Council through the provision of a postdoctoral fellowship (B.G.) and a PhD studentship (M.W.P.). We thank J. Raftery and M. Helliwell for X-ray crystallography.

Author Information X-ray crystallographic data (excluding structure factors) for **13** have been deposited at the Cambridge Crystallographic Data Centre, UK, under deposition number CCDC 701055. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk/data_request/cif). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.D.S. (john.sutherland@manchester.ac.uk).

LETTERS

Interior pathways of the North Atlantic meridional overturning circulation

Amy S. Bower[1], M. Susan Lozier[2], Stefan F. Gary[2] & Claus W. Böning[3]

To understand how our global climate will change in response to natural and anthropogenic forcing, it is essential to determine how quickly and by what pathways climate change signals are transported throughout the global ocean, a vast reservoir for heat and carbon dioxide. Labrador Sea Water (LSW), formed by open ocean convection in the subpolar North Atlantic, is a particularly sensitive indicator of climate change on interannual to decadal timescales[1–3]. Hydrographic observations made anywhere along the western boundary of the North Atlantic reveal a core of LSW at intermediate depths advected southward within the Deep Western Boundary Current (DWBC)[4–9]. These observations have led to the widely held view that the DWBC is the dominant pathway for the export of LSW from its formation site in the northern North Atlantic towards the Equator[10,11]. Here we show that most of the recently ventilated LSW entering the subtropics follows interior, not DWBC, pathways. The interior pathways are revealed by trajectories of subsurface RAFOS floats released during the period 2003–2005 that recorded once-daily temperature, pressure and acoustically determined position for two years, and by model-simulated 'e-floats' released in the subpolar DWBC. The evidence points to a few specific locations around the Grand Banks where LSW is most often injected into the interior. These results have implications for deep ocean ventilation and suggest that the interior subtropical gyre should not be ignored when considering the Atlantic meridional overturning circulation.

Profiling floats[12] released in the Labrador Sea during the 1990s showed little evidence of southward export of LSW in the DWBC[13–16]. This result was surprising because the DWBC is widely thought to be the dominant LSW export pathway towards the subtropics and tropics. Why did these floats not follow the DWBC into the subtropics? Were they biased by upper-ocean currents when they periodically ascended to the sea surface to fix their position, as recently suggested by numerical model results[17]? Were they released mainly in the recirculating waters of the subpolar gyre? Or is the DWBC in fact not the dominant export pathway for LSW?

To address these questions, 76 acoustically tracked Range and Fixing of Sound (RAFOS) floats[18], which do not need to surface to fix their position, were sequentially released in the DWBC near 50° N from 2003 to 2006 at two LSW depths, 700 and 1,500 m, for two-year drifting missions (see Fig. 1a and Methods for more details). Here we describe the spreading pathways of LSW revealed by the first 40 high-resolution RAFOS float trajectories, ten additional float displacement vectors and simulated trajectories (e-floats) from a high-resolution numerical ocean circulation model[19].

All RAFOS floats initially drifted southward in the DWBC after release at 50° N (Fig. 1b). But a large fraction of the floats—about 75% (29/40)—escaped from the DWBC before reaching the southern tip or 'Tail' of the Grand Banks (43° N) (Fig. 2a and b) and drifted into the interior. Many of these followed an eastward path along the subpolar–subtropical gyre boundary (Fig. 1a and b). Only 8% of all floats (3/40) followed the DWBC continuously from launch around the Tail of the Grand Banks. This is more than the number of profiling floats from the Labrador Sea that rounded the Tail of the Grand Banks in the DWBC (zero)[14], but is still a remarkably low number in light of the expectation that the DWBC is the dominant southward pathway for LSW.

A larger percentage of the RAFOS floats—about 23% (9/40)—reached the subtropics via an interior pathway, indicated by the cluster of trajectories extending south of 42° N in the longitude band 40°–60° W (Fig. 1b). The warmer temperatures measured by these floats indicate that they crossed the Gulf Stream into the subtropical gyre. The dominance of the interior versus DWBC pathway is further supported by the larger ensemble of 50 RAFOS float displacement vectors (Fig. 1b inset)—about 24% (12/50) surfaced south of 42° N in the interior (east of 60° W). Furthermore, the largest southward float displacements over two years were made by floats following an interior, not DWBC path (Fig. 1b inset). Interior pathways for the southward spreading of LSW into the subtropics have been suggested previously[7,9,17,20,21] but these float tracks offer the first evidence of the relative dominance of this pathway compared to the DWBC.

The RAFOS float trajectories reveal two primary locations where LSW escapes from the DWBC and enters the interior ocean—at the southeastern corner of Flemish Cap (especially for 1,500 m floats) and just upstream of the Tail of the Grand Banks (Fig. 2a and b). At these locations, the North Atlantic Current (Fig. 1a) is closest to the continental slope, supporting a previous conjecture that onshore excursions of the North Atlantic Current temporarily interrupt the flow of the DWBC and divert LSW into the interior[15].

To complement this analysis of the necessarily limited number of RAFOS float trajectories, simulated trajectories were generated using the eddy-resolving (~1/12°) primitive equation Family of Linked Atlantic Models Experiment (FLAME) model[19] (see Methods for details of trajectory computation). The e-float trajectories were calculated using the three-dimensional (x, y, z), time-varying model velocity fields to simulate fluid parcel motion as accurately as possible. The constant-pressure RAFOS floats drift only with the two-dimensional (x–y) flow field, but no significant differences were found in the model results using the two-dimensional or three-dimensional model velocity fields, in contrast to a previous modelling analysis of LSW pathways which used time-mean (as opposed to the time-varying fields used here) model velocity fields[17] (see Supplementary Information).

Seventy-two e-floats were initialized in the DWBC near 50° N with the same spatial and temporal pattern as the RAFOS floats. The spread of the model and RAFOS float trajectories after two years is very similar (Fig. 3a). There is little evidence for a continuous DWBC pathway; rather, e-floats tend to recirculate within the subpolar gyre and drift southward into the subtropical gyre interior. The loss of

[1]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02540, USA. [2]Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, North Carolina 27708, USA. [3]IFM-GEOMAR Leibniz-Institut für Meereswissenschaften, Kiel, 24105, Germany.

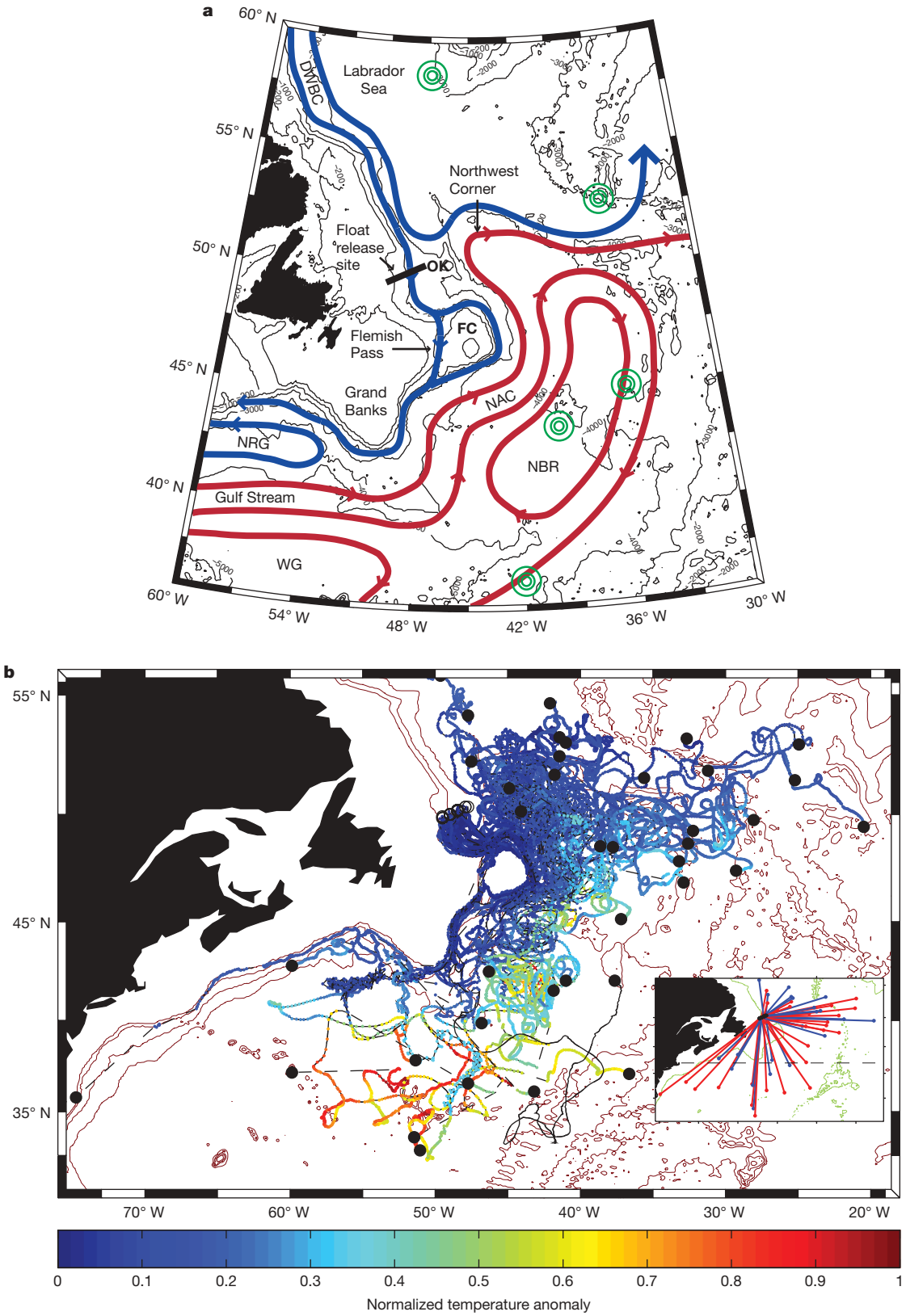


Figure 1 | Study area and RAFOS float trajectories at the LSW level in the western North Atlantic. **a**, Schematic diagram of the intermediate-depth circulation in the northwestern North Atlantic, with blue and red lines indicating cold and warm water pathways, respectively. Green concentric circles show locations of sound sources used to track floats. FC, Flemish Cap; NAC, North Atlantic Current; NBR, Newfoundland Basin Recirculation Gyre; NRG, Northern Recirculation Gyre; OK, Orphan Knoll; WG, Worthington Gyre. **b**, Two-year trajectories of 40 acoustically tracked RAFOS floats released at 700 and 1,500 m in the DWBC near 50° N. Positions are indicated daily with colour-coded dots, where the colour indicates the normalized temperature anomaly, defined as $(T - T_i)/\delta T_{max}$. T_i is each float's initial temperature, and δT_{max} is the maximum temperature range observed by the floats as a group, 6.4 °C at 700 dbar and 1.8 °C at 1,500 dbar. Dashed lines indicate missing track. The inset shows the two-year displacement vectors for the same floats plus ten more that have yet to be processed, colour-coded by depth (red for 700 m and blue for 1,500 m).

e-floats from the DWBC is also very similar to that observed with the RAFOS floats (Fig. 2b).

This favourable comparison supports extending the integration to generate longer simulated trajectories (Fig. 3b and c), beyond the technical capabilities of the RAFOS floats. After five years, the Tail of the Grand Banks begins to stand out as a barrier to the westward spread of e-floats in the DWBC. Only after ten years is a thin collection of a small number of trajectories evident within the DWBC west of the Grand Banks, emphasizing the importance of recirculation in the Newfoundland Basin in slowing the equatorward transport of recently ventilated LSW in the DWBC[7,22].

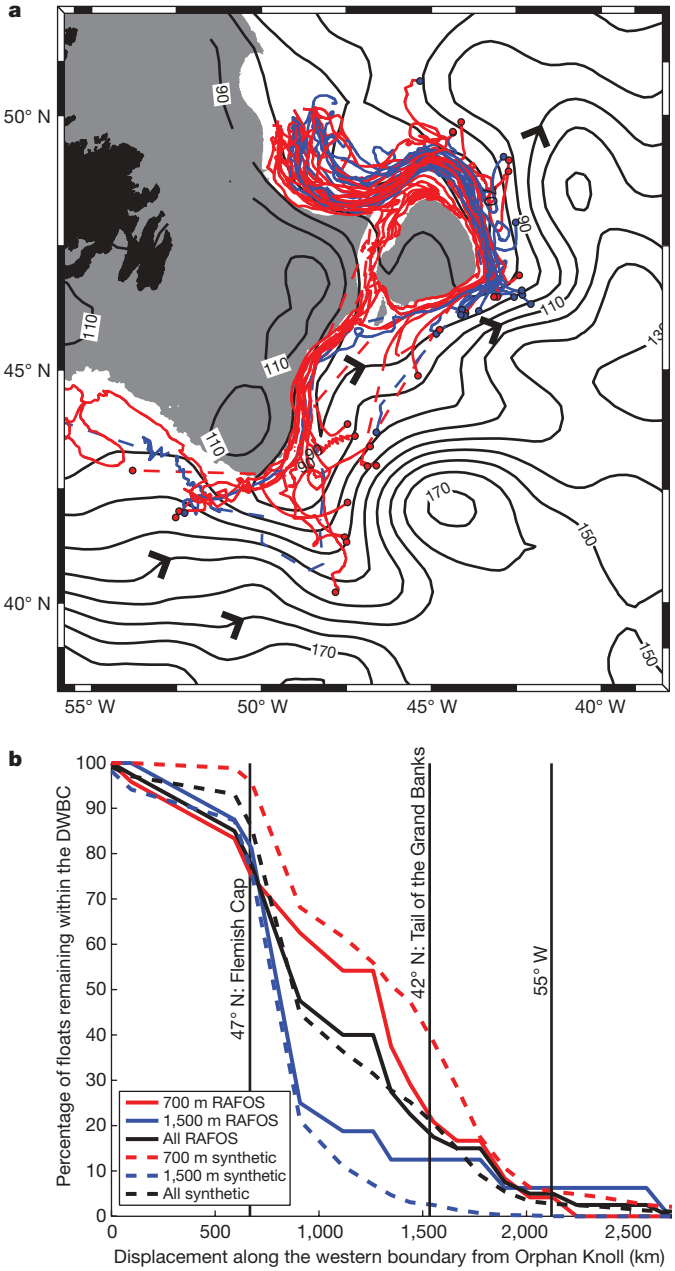


Figure 2 | Loss of floats from the DWBC. a, Trajectories of 40 RAFOS floats (blue, 1,500 m; red, 700 m) between launch position and the position where they first cross the 4,000 m isobath (coloured dots) illustrate where floats were most likely to leave the DWBC and drift into the interior. The mean path of the Gulf Stream and North Atlantic Current is shown with the mean absolute dynamic topography from Aviso (Archiving, Validation and Interpretation of Satellite Oceanographic data) for the float sampling time period. Arrows indicate direction of geostrophic surface flow, and the gradient is proportional to flow speed. The path of the North Atlantic Current is similar to that derived from subsurface floats[26] and hydrographic data[27]. The 700-m isobath is shaded grey. **b**, Retention of RAFOS floats (solid lines) and e-floats (dashed lines) in the DWBC as a function of along-boundary distance from 50° N.

To quantify the Lagrangian spreading pathways of LSW, 7,280 e-floats were released and integrated for 15 years and a two-dimensional histogram of float position was mapped (Fig. 3d) (see Supplementary Information for details of map construction). The sharp drop in e-float concentration around the Grand Banks, and the southward penetration into the subtropical interior are clearly revealed. The e-floats are concentrated within an eddy-driven circulation that has previously been postulated to provide interior pathways from subpolar to subtropical latitudes[20,21].

A further demonstration of the lack of strong connectivity of LSW pathways around the Grand Banks is given by 15-year back trajectories for e-floats that arrived at Line W (~69° W), where the properties and transport of the subtropical DWBC are being monitored (see http://www.whoi.edu/science/PO/linew) (Fig. 3e). Again, a strong discontinuity appears at the Tail of the Grand Banks. A thin ribbon of trajectories is traced from the Tail of the Grand Banks upstream to the western boundary of the Labrador Sea, but represents only a small fraction of the total at Line W. The model DWBC in the subtropical basin is mainly transporting waters that are recirculating north of the Gulf Stream and west of the Grand Banks in the Northern Recirculation Gyre (Fig. 1a)[23].

To quantify the relative importance of the DWBC versus interior pathways in the model, we mapped the transport associated with e-floats that drifted from the float release site at 50° N to 32° N within 15 years (Fig. 4; see Supplementary Information for details of map construction). We kept track of the e-floats that (1) never crossed offshore of the 4,000 m isobath into the interior (exclusively inshore), (2) were inshore of the 4,000 m isobath but may have crossed that isobath at some point (all inshore) and (3) were offshore of the 4,000 m isobath (all offshore). Transport values for each group as a function of distance along the boundary are tabulated in the Supplementary Information.

At the release site, all transport is inshore of the 4,000 m isobath. Moving southward along the path of the DWBC to the Tail of the Grand Banks, the all-inshore transport drops to about 62%, and the exclusively-inshore transport drops even more (43%). The transport located in the interior grows accordingly. A similar result for the all-inshore transport at the Tail of the Grand Banks was obtained in a previous modelling study[17], from which the authors concluded that the DWBC was the dominant pathway for the export of LSW. However, as seen in Fig. 4, the all-inshore and especially the exclusively-inshore transports drop precipitously moving around the southern tip of the Grand Banks—at 55° W the all-inshore and exclusively-inshore transports are only 11.5% and 2.6%, respectively. At Cape Hatteras (36° N), only 3.1% of the transport being tracked is located inshore and 0.1% followed the DWBC continuously from the release site. South of 34° N, the interior transport begins to converge back towards the western boundary, but clearly the vast majority of the LSW transport tagged at 50° N in the DWBC that reached 32° N did so via an interior pathway. This result is consistent with the relatively larger number of RAFOS floats entering the subtropical gyre interior south of the Grand Banks (Fig. 1b) and with the observation of relatively young tracer ages there[7].

The directions of LSW spreading presented here are generally consistent with those inferred from hydrographic and tracer studies: eastward and northward within the subpolar gyre, into the subtropical interior and along the DWBC[4,7,9,24]. However, the new float observations and simulated float trajectories provide evidence that the southward interior pathway is more important for the transport of LSW through the subtropics than the DWBC, contrary to previous thinking. Though the DWBC is easier to observe—a well-defined, relatively stationary current close to shore compared to the vast, turbulent and unconstrained interior—our results suggest that further study of the interior subtropical gyre and the complex region around the Grand

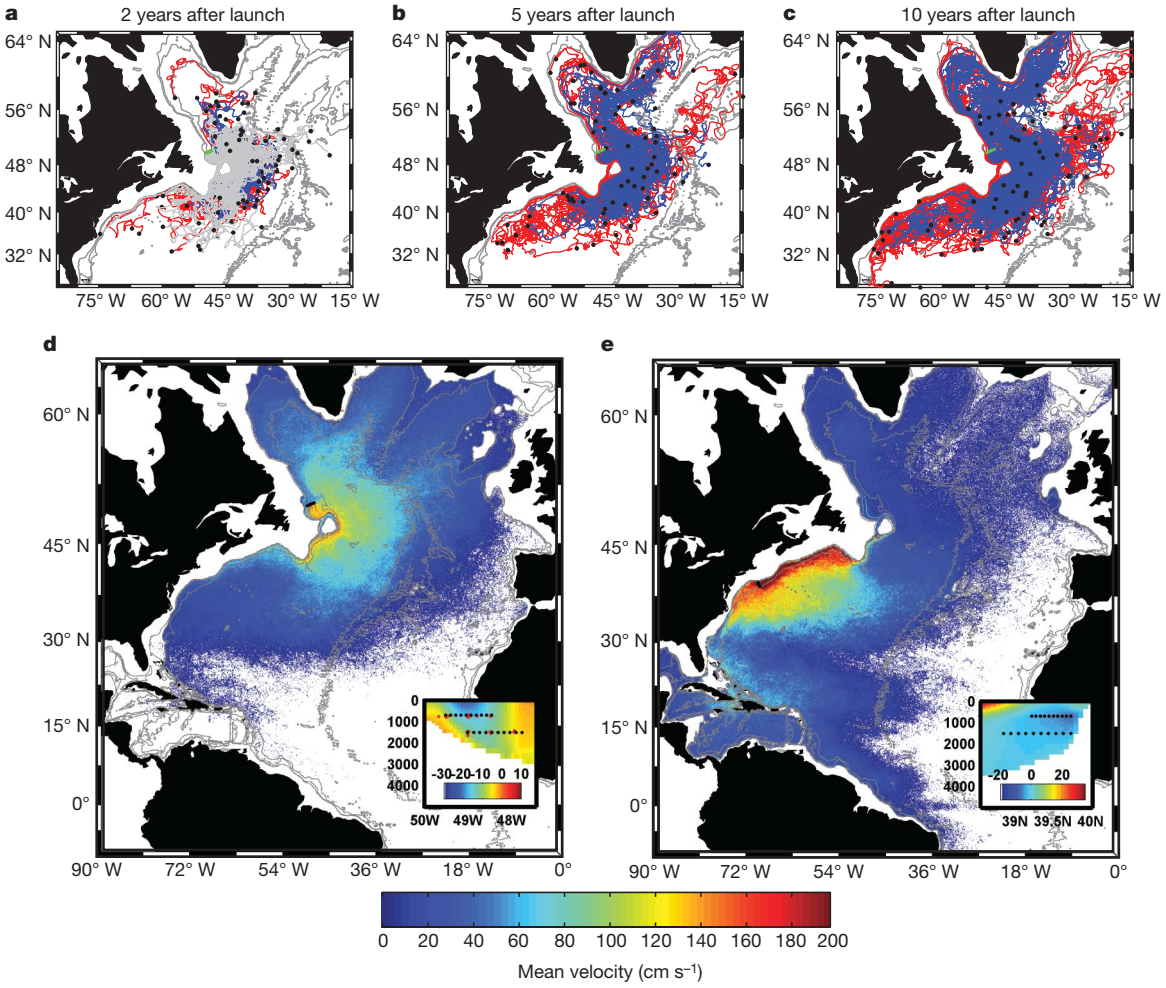


Figure 3 | Simulated trajectories from FLAME. Trajectories of (**a**) 2, (**b**) 5 and (**c**) 10 years for 72 e-floats at 700 m (red) and 1,500 m (blue), selected from an ensemble of 7,280 15-year trajectories initiated at the RAFOS float release sites near 50 °N. The model trajectories were computed using the three-dimensional model velocity fields, so the virtual particles change their depth accordingly. The RAFOS trajectories (light grey) are shown in **a** for comparison. The endpoint of each e-float trajectory is marked with a black dot. Isobaths are shown in darker grey for 0, 700, 1,500 and 3,000 m. **d** and **e**, 7,280 forward e-trajectories launched at Orphan Knoll (**d**) and 7280 backward trajectories launched at Line W (**e**) in the core of the DWBC, condensed into float location two-dimensional histogram maps. The float launch locations are shown in black. The insets to each map show the float launch locations at each site superposed on the mean velocity (in cm s^{-1}) cross-section from the FLAME model. The RAFOS and e-float launch points are shown with red and black dots, respectively.

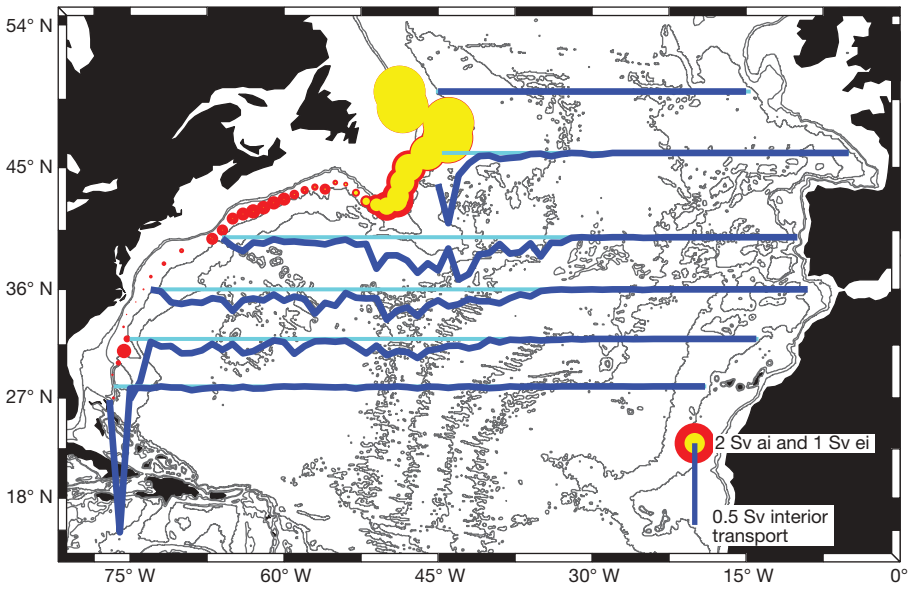


Figure 4 | Transport map for 1,338 e-floats released in the layer 703–1,548 m at 50° N that crossed the latitude 32° N within 15 years. Coloured circles indicate transport associated with all-inshore ('ai', red) and exclusively-inshore ('ei', yellow) e-floats, where circle radius is proportional to transport in Sv (see scale bar). Blue lines indicate transport associated with all-offshore ('ao') e-floats as a function of longitude for selected latitudes. Light blue lines show the zero reference for each blue line. Details on map construction as well as tabulated transport values are given in the Supplementary Information.

Banks is needed to understand better the pathways of the deep limb of the Atlantic Meridional Overturning Circulation.

METHODS SUMMARY

The RAFOS floats used in this study were released (nominally) in groups of six every three months between July 2003 and April 2005 along a section extending from Cape Bonavista, Newfoundland to Orphan Knoll. As the floats drifted, their positions were determined relative to sound sources moored in the eastern and western North Atlantic. The floats were isobaric (constant pressure) and ballasted to drift at two levels corresponding to the tracer cores of Upper LSW (700 dbar) and Classical LSW (1500 dbar). The floats internally recorded travel times from the sound sources, as well as temperature and pressure measurements once daily for two years, before returning to the surface and transmitting all the collected data via the Argos satellite-based data retrieval system.

The simulated trajectories presented in this study were generated using the FLAME model. This model was based on the MOM2.1 code[25] and modified as part of the FLAME project[4]. Following a ten-year spin-up from rest with climatological forcing, this model was run with interannually varying wind stresses and heat fluxes for the period 1987–2004. The model output consists of three-dimensional snapshots of horizontal velocity, temperature and salinity fields over the domain on a 1/12° resolution Mercator grid.

To calculate the simulated trajectories, model velocity fields from the years 1994, 1996 and 1998 were repeated sequentially for 15 years. These years represent a variety of forcing states as indicated by the North Atlantic Oscillation index. Model floats were initialized sequentially over the course of the first three years and every trajectory was computed for 15 years using 3-day snapshot, three-dimensional velocity fields. Thus the virtual floats are displaced both horizontally and vertically in accordance with the velocity fields to simulate water parcel movement as accurately as possible.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April 2008; accepted 5 March 2009.

1. Lazier, J. Oceanographic conditions at ocean weather ship Bravo, 1964–1974. *Atmos. Ocean* **18,** 227–238 (1980).
2. Rhines, P. B. & Lazier, J. R. N. A 13-year record of convection and climate change in the deep Labrador Sea. In *Abstract Report of ACCP Principal Investigators Meeting* (NOAA Office of Global Programs, 1995).
3. Lazier, J., Yashayaev, I., Rhines, P., Hendry, R. & Clarke, A. Convection and restratification in the Labrador Sea, 1990–2000. *Deep-Sea Res. I* **49,** 1819–1835 (2002).
4. Talley, L. D. & McCartney, M. S. Distribution and circulation of Labrador Sea Water. *J. Phys. Oceanogr.* **12,** 1189–1205 (1982).
5. Molinari, R. L., Fine, R. A. & Johns, E. The Deep Western Boundary Current in the tropical North Atlantic Ocean. *Deep-Sea Res.* **39,** 1967–1984 (1992).
6. Pickart, R. S. Water mass components of the North Atlantic Deep Western Boundary Current. *Deep-Sea Res.* **39,** 1553–1572 (1992).
7. Smethie, W. M. Jr, Fine, R. A., Putzka, A. & Jones, E. P. Tracing the flow of North Atlantic Deep Water using chlorofluorocarbons. *J. Geophys. Res.* **105,** 14297–14324 (2000).
8. Stramma, L. *et al.* Deep water changes at the western boundary of the subpolar North Atlantic during 1996 to 2001. *Deep-Sea Res. I* **51,** 1033–1056 (2004).
9. Rhein, M. *et al.* Labrador Sea Water: pathways, CFC inventory, and formation rates. *J. Phys. Oceanogr.* **32,** 648–665 (2002).
10. Molinari, R. L. *et al.* The arrival of recently formed Labrador Sea Water in the Deep Western Boundary Current at 26.5° N. *Geophys. Res. Lett.* **25,** 2249–2252 (1998).
11. Bryden, H. L., Johns, W. E. & Saunders, P. M. Deep Western Boundary Current east of Abaco: Mean structure and transport. *J. Mar. Res.* **63,** 35–57 (2005).
12. Davis, R. E. The Autonomous Lagrangian Circulation Explorer (ALACE). *J. Atmos. Ocean. Technol.* **9,** 264–285 (1992).
13. Lavender, K. L., Davis, R. E. & Owens, W. B. Mid-depth recirculation observed in the interior Labrador and Irminger seas by direct velocity measurements. *Nature* **407,** 66–69 (2000).
14. Lavender, K. L., Owens, W. B. & Davis, R. E. The mid-depth circulation of the subpolar North Atlantic Ocean as measured by subsurface floats. *Deep Sea Res. I* **52,** 767–785 (2005).
15. Fischer, J. & Schott, F. A. Labrador Sea Water tracked by profiling floats—from the Boundary Current into the open North Atlantic. *J. Phys. Oceanogr.* **32,** 573–584 (2002).
16. Schott, F. A. & Brandt, P. in *Ocean Circulation: Mechanisms and Impacts* (eds Schmittner, A., Chiang, J. C. H. & Hemming, S.) 95 (American Geophysical Union, 2007).
17. Getzlaff, K., Boening, C. W. & Dengg, J. Lagrangian perspectives of deep water export from the subpolar North Atlantic. *Geophys. Res. Lett.* **33,** L21S08, doi:10.1029/2006GL026470 (2006).
18. Rossby, H. T., Dorson, D. & Fontaine, J. The RAFOS system. *J. Atmos. Ocean. Technol.* **3,** 672–679 (1986).
19. Boening, C. W., Scheinert, M., Dengg, J., Biastoch, A. & Funk, A. Decadal variability of subpolar gyre transport and its reverberation in the North Atlantic overturning. *Geophys. Res. Lett.* **33,** L21S01 (2006).
20. Lozier, M. S. Evidence for large-scale eddy-driven gyres in the North Atlantic. *Science* **277,** 361–364 (1997).
21. Lozier, M. S. The impact of mid-depth recirculations on the distribution of tracers in the North Atlantic. *Geophys. Res. Lett.* **26,** 219–222 (1999).
22. Schmitz, W. J. & McCartney, M. S. On the North Atlantic Circulation. *Rev. Geophys.* **31,** 29–49 (1993).
23. Hogg, N. G., Pickart, R. S., Hendry, R. M. & Smethie, W. J. Jr. The northern recirculation gyre of the Gulf Stream. *Deep-Sea Res.* **33,** 1139–1165 (1986).
24. Kieke, D. *et al.* Changes in the CFC inventories and formation rates of Upper Labrador Sea Water, 1997–2001. *J. Phys. Oceanogr.* **36,** 64–86 (2006).
25. Pacanowski, R. C. *MOM 2 (Version 2.0) Documentation User's Guide and Reference Manual*, Geophys. Fluid Dynam. Lab. Ocean Technol. Rep. 3.1 329 (GFDL/NOAA, 1996).
26. Rossby, H. T. The North Atlantic Current and surrounding waters: at the crossroads. *Rev. Geophys.* **34,** 463–481 (1996).
27. Kearns, E. J. & Rossby, H. T. Historical position of the North Atlantic Current. *J. Geophys. Res. C* **103,** 15509–15524 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the captains and crews of the R/V *Oceanus* and numerous Canadian fisheries research vessels, and scientists from the Northwest Atlantic Fisheries Centre in St. John's, Newfoundland, for their assistance in the deployment of the floats and sound sources used in this study. In particular, we thank E. Colbourne for his support of the float release programme. We also acknowledge J. Valdes and B. Guest for their expert technical support in the preparation of the floats, and H. Furey for providing float tracking. J. Fischer of IFM-GEOMAR provided several days of ship time to replace sound sources. We dedicate this work to the memory of F. Schott. The work was supported by the US National Science Foundation.

Author Contributions A.S.B. and M.S.L. contributed equally to this work. A.S.B. led the RAFOS float field programme, and analysed the float and altimetric data. M.S.L. led the modelling study and with S.F.G. analysed the simulated trajectories. C.W.B. provided the model output and assisted with the calculation of the simulated trajectories from FLAME.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S.B. (abower@whoi.edu) and M.S.L. (mslozier@duke.edu).

METHODS

RAFOS floats. The RAFOS floats used in this study were released (nominally) in groups of six floats every three months between July 2003 and April 2005 along a section extending from Cape Bonavista, Newfoundland to Orphan Knoll, in water depths between 1,400 and 2,800 m (see Supplementary Table S1 for details). As the floats drifted, their positions were determined relative to sound sources moored in the eastern and western North Atlantic. All but the first float setting were made from various Canadian research vessels by the Northwest Atlantic Fisheries Centre in St. John's, Newfoundland, during spring, summer and autumn cruises. To release RAFOS floats during winter, six dual-release floats were deployed during each autumn cruise in addition to the six regular floats. The dual-release floats each had a heavy length of chain attached that initially anchored them to the sea floor, creating a 'float park'[28]. These floats were programmed to release the anchor chain on the following February 15th, and then drift to their ballast depth to begin their two-year mission.

The RAFOS floats used in this study were isobaric and ballasted to drift at two levels, corresponding to Upper LSW (700 dbar) and Classical LSW (1,500 dbar). The floats collected position, temperature and pressure information once daily for two years, then returned to the surface to transmit all the collected data via Service ARGOS.

Satellite altimetry. In Fig. 2a, the path of the Gulf Stream and North Atlantic Current were determined using maps of absolute dynamic topography produced by Ssalto/Duacs at Collecte Localization Satellites, a subsidiary of the French Space Agency (CNES) and the French Research Institute for Exploration of the Sea (IFREMER). This product is generated using all available satellite missions since 1992. With support from CNES it is distributed online by Aviso (http://www.jason.oceanobs.com/html/donnees/produits/hauteurs/global/madt_uk.html). The maps of absolute dynamic topography combine gridded (1/3°) sea level anomaly fields with the Combined Mean Dynamic Topography (Rio05)[29].

Synthetic float trajectory calculations. The synthetic trajectories used in this study were generated using the FLAME model, which was based on the MOM2.1 code and modified as part of the FLAME project[19]. Following a ten-year spin-up from rest with climatological forcing, this model was run with interannually varying wind stresses and heat fluxes for the period 1987–2004. Model output consists of three-dimensional snapshots of horizontal velocity, temperature and salinity fields over the domain on a 1/12° resolution Mercator grid. In the vertical, the domain was split into 45 *z*-coordinate levels. The vertical velocity was computed from the horizontal velocity by requiring that the local divergence of the three-dimensional velocity field be zero throughout the model domain.

Velocity fields from FLAME model years 1994, 1996 and 1998, repeated sequentially, were used for the 15-year trajectories. These years represent a variety of forcing states as indicated by the North Atlantic Oscillation index. The e-floats were released sequentially over the course of the first three years and every trajectory was computed for 15 years using 3-day snapshot three-dimensional model velocity fields. We note that throughout the study, '700 m' and '1,500 m' e-floats refer to their approximate depths of float initialization. Subsequent e-float positions are estimated from the three-dimensional model velocity fields, so the virtual floats are displaced both horizontally and vertically to simulate water parcel movement as accurately as possible.

Computation of float loss from the DWBC. See Fig. 2b. Because the DWBC generally flows inshore of the 4,000 m isobath in the study region, a RAFOS or e-float was considered out of the boundary current if it crossed this isobath into deeper water. To determine the number of floats that remain within the DWBC at different points along the coast, ten boxes, each spanning the width of the continental slope, were defined along the boundary. The number of floats that passed through each box was counted. In this analysis, floats that left the DWBC at any point along the boundary were never counted again, even if they happened to re-enter one of the boxes. Thus, the number of floats remaining within the DWBC includes only those floats that have remained in the DWBC continuously since launch (also called exclusively-inshore floats).

Construction of e-float position histograms. See Fig. 3d and e. To present the Lagrangian pathway information from the thousands of synthetic trajectories used in this study efficiently, a two-dimensional histogram of float positions, essentially a map of float concentration, was used. A count was made of the number of floats that passed through each 1/12° horizontal bin; repetitions of the same float were counted. The units on the two-dimensional histogram are the number of floats passing through each bin. Histograms of the 700 and 1,500 m subsets of the float population are qualitatively similar to the whole population histograms.

Construction of transport map. See Fig. 4. The e-floats were initialized at the RAFOS float release site (near 50° N) in the layer spanning 703 to 1,540 m. The e-floats were launched on a 7-level grid with nodes at: 744, 828, 920, 1,022, 1,140, 1,280 and 1,448 m. Each e-float was assigned a transport computed from the velocity, layer thickness and cell width at the e-float's release location. The layer thicknesses used to compute the transport tag for each float range from 78–184 m, increasing with increasing layer depth. The three-dimensional trajectories were computed using the repeating cycle of 1994, 1996 and 1998 3-day updated velocity fields for 15-year integrations. At release, the total transport was 12 Sv in the layer (for each of the 36 launch dates) divided between a grand total of 6,539 floats. Floats were launched every 30 days for the first three years (and because the velocity field repeated itself after the first three years, no new initializations were made after that point). Only those e-floats that crossed 32° N within 15 years were retained, which accounts for the movement of 2 Sv (average transport per launch initialization) among a total of 1,338 e-floats. Longer integrations gave very similar results.

28. Zenk, W., Pinck, A., Becker, S. & Tillier, P. The Float Park: A new tool for a cost-effective collection of Lagrangian time series with dual release RAFOS floats. *J. Atmos. Ocean. Technol.* **17,** 1439–1443 (2000).
29. Rio, M. Schaeffer, P., Lemoine, J. & Hernandez, F. Estimation of the ocean Mean Dynamic Topography through the combination of altimetric data, in-situ measurements and GRACE geoid: from global to regional studies. In *Proc. of the GOCINA Int. Workshop* (GOCINA, 2005).

Vol 459|14 May 2009|doi:10.1038/nature07995

LETTERS

A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany

Nicholas J. Conard[1]

Despite well over 100 years of research and debate, the origins of art remain contentious[1–3]. In recent years, abstract depictions have been documented at southern African sites dating to ~75 kyr before present (BP)[4,5], and the earliest figurative art, which is often seen as an important proxy for advanced symbolic communication, has been documented in Europe as dating to between 30 and 40 kyr BP[2]. Here I report the discovery of a female mammoth-ivory figurine in the basal Aurignacian deposit at Hohle Fels Cave in the Swabian Jura of southwestern Germany during excavations in 2008. This figurine was produced at least 35,000 calendar years ago, making it one of the oldest known examples of figurative art. This discovery predates the well-known Venuses from the Gravettian culture by at least 5,000 years and radically changes our views of the context and meaning of the earliest Palaeolithic art.

Excavators recovered the six fragments of carved ivory that form the Venus (Fig. 1) between 8 and 15 September 2008. The importance of the discovery became apparent on 9 September when the main piece of the sculpture, which represents the majority of the torso, was recovered. Two of the fragments were documented *in situ* and measured in three dimensions. Four fragments were recovered in connection with water screening and can be localized to a 10-l volume corresponding to a ~3-cm-thick portion of a quarter metre. The pieces of the figurine lay about 3 m below the current surface of the cave in an area about 20 m from the cave's entrance. All of the finds come from the southwest quadrant of a single square metre and were recovered from within 12 cm in the vertical dimension (Fig. 2). Although, owing to their fragility and complex depositional histories, many of the ivory artworks from the Swabian Jura are highly fragmentary, the Venus from Hohle Fels is nearly complete; only the left arm and shoulder are missing. The excellent preservation and the close stratigraphic association of the pieces of the figurine indicate that the Venus experienced little taphonomic disturbance after deposition. The quarter metre in which the figurine was found borders directly on the western edge of the dig, raising the possibility that the missing portion may be recovered as excavation continues.

The figurine originates from a red-brown, clayey silt at the base of ~1 m of Aurignacian deposits. One fragment was attributed to feature 10, a small area rich in charcoal at the base of archaeological

Figure 1 | Side and front views of the Venus of Hohle Fels. Photos by H. Jensen; copyright, University of Tübingen.

[1]Abteilung für Ältere Urgeschichte und Quartärökologie, Institut für Ur- und Frühgeschichte und Archäologie des Mittelalters, Universität Tübingen, Schloss Hohentübingen, 72070 Tübingen, Germany.

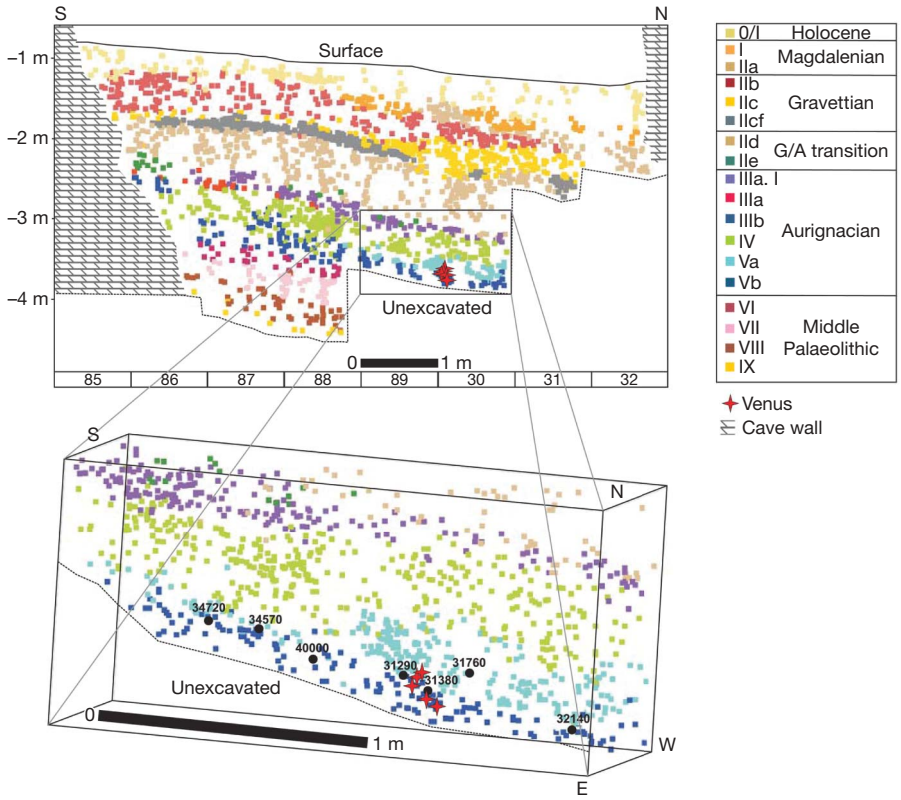


Figure 2 | Stratigraphic position of the Venus of Hohle Fels and associated radiocarbon dates from archaeological horizon Va feature 10 and Vb. The lower plot shows the radiocarbon dates (in years before present) of some of the samples found near the Venus (Table 1). Figure by M. Malina.

horizon Va, directly overlying archaeological horizon Vb. The remaining five pieces were recovered from archaeological horizon Vb, which is an approximately 8-cm-thick deposit of clayey silt directly overlying the sterile clays that separate the Aurignacian from the underlying Middle Palaeolithic strata. The Venus lay in pieces next to a number of limestone blocks with dimensions of several decimetres. The find density in this part of archaeological horizon Vb is moderately high, with much flint-knapping debris, worked bone and ivory, faunal remains of horse, reindeer, cave bear, mammoth and ibex, and burnt bone.

Six new radiocarbon measurements on bone and one on charcoal from feature 10 and archaeological horizon Vb have been made at the Oxford Radiocarbon Accelerator Unit (Table 1). Four of the dates fall between 31.3 and 32.1 kyr BP. Two other dates fall in the range 34.6–34.7 kyr BP. One bone dates from 40.0 kyr BP. The new series of dates on bones from the vicinity of the Venus were all made on collagen processed using ultrafiltration[6]. The amount of collagen ranged from 2.2 to 11.4% in the six bones sampled. Two additional measurements on bone and one on charcoal from the 2002 excavation were made at the Leibniz Laboratory, Kiel, and yielded dates between 33.3 and 35.7 kyr BP. These finds come from the same stratigraphic position 2 m farther to the southeast. The samples from the 2002 excavation were initially classified as belonging to archaeological horizon Va, but on stratigraphic grounds have been redesignated as belonging to archaeological horizon Vb. Five dates of bones recovered during the 2007 excavation from archaeological horizon Va, in a find-rich wedge of sediment between archaeological horizons IV and Vb, were measured in Kiel and fall in the range 31.7–32.3 kyr BP[7,8]. Previously, a sculpture of a waterfowl and a therianthrop were recovered from archaeological horizon IV, where nine radiocarbon dates measured in Kiel and Oxford on bone fall between 30 and 33 kyr BP[9]. All of the bones measured in Kiel were well preserved and yielded between 6.4 and 18.6% collagen. Most of the bones dated at Kiel and Oxford show anthropogenic modifications, and the two pieces of charcoal from archaeological horizon Vb almost certainly originate from anthropogenic fires.

This wide range of dates from archaeological horizon Vb presents a situation similar to that from the nearby site of Geißenklösterle, where the lower Aurignacian deposit of archaeological horizon III has produced 33 radiocarbon dates between 29 and 40 kyr BP[8]. The same horizon has yielded thermoluminescence dates in the range of 40 kyr BP[10].

There is no simple explanation for the variable radiocarbon dates from Hohle Fels and Geißenklösterle. The noisy signals result from a combination of factors including variable sample preparation, variable levels of atmospheric carbon, taphonomic mixing and excavation error[8,11,12]. Given the lack of reproducibility within and between radiocarbon laboratories, I prefer to emphasize the stratigraphic context of the finds, and to use the highly variable radiometric dates as rough indicators of age[8]. Although there is no generally accepted calibration for radiocarbon dates over 30 kyr BP, preliminary calibrations suggest that dates of 32 kyr BP correspond to roughly 36 kyr BP in calendar years[13]. If the early dates are correct, the Venus would be even older. The fact that the Venus is overlain by five Aurignacian horizons, containing a dozen stratigraphically intact anthropogenic features with a total thickness of ~1 m, suggests that the figurine is of an age corresponding to the start of the Aurignacian, around 40,000 calendar years ago. The overlying deposits contain rich assemblages of Aurignacian lithics, organic tools and personal ornaments, as well as three examples of figurative art[14]. We do not have reliable data on rates of sedimentation and the exact duration of the Aurignacian; however, Hohle Fels is one of the largest and most visible caves in the Swabian Jura, suggesting that it would be quickly occupied by the first Upper Palaeolithic people in the region.

Although much ivory-working debris has been recovered from the basal Aurignacian deposits at Geißenklösterle and Hohle Fels, this sculpture is the first example of figurative art recovered from the basal Aurignacian in Swabia. Unless scenarios involving major taphonomic disturbances and mixing with overlying sediments are considered, the discovery of the Venus of Hohle Fels refutes claims

Table 1 | AMS radiocarbon dates from the Aurignacian and Middle Palaeolithic of Hohle Fels

Laboratory number	Arch. horizon	Material	Modification	Collagen (%)	Date (years BP)	Cultural group
OxA-4979	III	*Salix* charcoal	—	—	27,600 ± 800	A?
KIA 32056	IIIa. 1	Reindeer metatarsal	Impact	8.0	$29{,}710^{+210}_{-200}$	A
KIA 32055	IIIa. 1	Cave bear rib	Cut mark	6.6	$30{,}340^{+290}_{-280}$	A
KIA 16038	IIIa	Reindeer femur	Impact + cut marks	14.4	29,840 ± 210	A
KIA 18877	IIIa	*Pinus* charcoal	—	—	$30{,}170^{+250}_{-240}$	A
OxA-4601	IIIa	Bone	—	—	30,550 ± 550	A
KIA 18876	IIIa	*Pinus* charcoal	—	—	$31{,}010^{+600}_{-560}$	A
KIA 16039	IIIa	Ungulate tibia	Impact	15.7	$31{,}140^{+250}_{-240}$	A
KIA 18878	IIIb	*Pinus* charcoal	—	—	$29{,}780^{+330}_{-310}$	A
KIA 3505	IIIb	Mammoth/rhino bone	Impact	—	$29{,}990^{+340}_{-330}$	A
OxA-4980	IV	*Salix* + *Betula* charcoal	—	—	28,750 ± 750	A
KIA 32057	IV	Reindeer radius/ulna	Impact	9.5	30,040 ± 210	A
KIA 32060	IV. 6	Long-bone fragment	Tool (retoucher)	6.6	$30{,}110^{+220}_{-210}$	A
KIA 32058	IV. 6	Horse mandible	Impact	3.7	30,420 ± 220	A
KIA 32059	IV. 6	Rib fragment	Tool (chisel)	7.1	$30{,}460^{+250}_{-240}$	A
OxA-4600	IV	Reindeer metapodial	—	—	31,100 ± 600	A
KIA 18879	IV	Unidentified charcoal	—	—	$31{,}160^{+1{,}530}_{-1{,}280}$	A
KIA 16037	IV	Reindeer/ chamois humerus	Impact + cut mark	14.4	$32{,}470^{+290}_{-280}$	A
KIA 16036	IV	Horse femur	Tool (retoucher)	15.5	$33{,}090^{+260}_{-250}$	A
KIA 35464	Va	Horse tibia/ radius	Tool (retoucher)	9.2	31,750 ± 260	A
KIA 35463	Va	Horse rib	Cut mark	14.2	$32{,}030^{+280}_{-270}$	A
KIA 35462	Va	Reindeer vertebra	Cut mark	9.5	$32{,}090^{+350}_{-340}$	A
KIA 35460	Va	Mammoth vertebra	—	6.4	$32{,}370^{+280}_{-270}$	A
KIA 35459	Va	Horse radius	Tool (retoucher)	10.3	$32{,}550^{+300}_{-290}$	A
OxA-19783*	Va. 10	Reindeer tibia	Cut mark	3.8	31,760 ± 200	A
OxA-19859*	Va. 10	Mammoth/ rhino rib	Impact	4.5	34,570 ± 260	A
OxA-19860*	Vb	*Pinus* charcoal	—	—	31,290 ± 180	A
OxA-19780*	Vb	Horse rib	Cut mark	11.4	31,380 ± 180	A
OxA-19779*	Vb	Horse tibia	Tool (retoucher)	3.6	34,720 ± 280	A
OxA-19782*	Vb	Horse hyoid	Cut mark	2.2	32,140 ± 310	A
KIA 16035	Vb**	Horse bone	Impact	17.8	33,290 ± 270	A
KIA 18880	Vb**	*Pinus* charcoal	—	—	$34{,}190^{+340}_{-330}$	A
KIA 16034	Vb**	Ungulate humerus	Impact + cut marks	18.6	$35{,}710^{+360}_{-340}$	A
OxA-19781*	Vb	Ibex tibia	Impact	4.1	40,000 ± 500	A
KIA 19564	VIb	Red deer metacarpal	Impact + cut marks	16.0	$35{,}760^{+660}_{-610}$	MP
KIA 19562	VIb	Cave bear metapod.	Possible cut mark	17.7	$36{,}380^{+380}_{-360}$	MP
KIA 19563	VII	Ibex/reindeer bone	Impact	12.8	$36{,}350^{+540}_{-510}$	MP
KIA 32054	VII	Cave bear rib	Possible cut mark	6.8	$37{,}940^{+530}_{-500}$	MP
KIA 32052	VIII	Reindeer tibia	Probable cut mark	2.9	$39{,}580^{+600}_{-560}$	MP
KIA 32053	IX	Bone	Impact	4.7	$38{,}560^{+530}_{-500}$	MP

AMS, accelerator mass spectrometry; A, Aurignacian; MP, Middle Palaeolithic.
* Previously unpublished dates.
** Originally published as archaeological horizon Va and changed to Vb on the basis of new stratigraphic observations. See ref. 8 for original publication of dates.

that figurative representations and other symbolic artefacts first appear in the later phases of the Swabian Aurignacian[11,12].

The Venus shows a range of entirely unique features as well as a number of characteristics present in later female figurines (Figs 1 and 3). Because carvings in mammoth ivory record many details, numerous specific observations can be made that allow comparisons with other Palaeolithic artworks. The vertical axis of the Venus runs parallel to the long axis of the mammoth tusk. The structure of the ivory shows that the two legs are oriented towards the proximal end of the tusk and the shoulders towards the distal end. The preserved portion of the figurine has a length of 59.7 mm, a width of 34.6 mm, a thickness of 31.3 mm and weighs 33.3 g.

The Venus of Hohle Fels lacks a head. Instead, an off-centre, but carefully carved, ring is located above the broad shoulders of the figurine. This ring, despite being weathered, preserves polish, suggesting that the figurine at times was suspended as a pendant. The shape of the preserved part of the figurine is asymmetrical, with the right shoulder elevated above the left side of the figurine. Beneath the shoulders, which are roughly as thick as they are wide, large breasts project forwards. The figurine has two short arms with two carefully carved hands resting on the upper part of the stomach below the breasts. Each hand has precisely carved fingers, with five clearly visible on the left hand and four on the right hand. The navel is visible and correctly placed anatomically.

The Venus has a short, squat form with a waist slightly narrower than the broad shoulders and wide hips. Multiple, deeply incised horizontal lines cover the abdomen from the area below the breasts to the pubic triangle. Several of these horizontal lines extend to the back of the figurine and are suggestive of clothing or a wrap of some kind. Microscopic images show that these incisions were created by repeatedly cutting along the same lines with sharp stone tools (Fig. 3). Such deep cuts into ivory are only possible with the application of significant force.

The legs of the Venus are short, pointed and asymmetrical, with the left leg noticeably shorter than the right leg. The buttocks and genitals are depicted in more detail. The split between the two halves of the buttocks is deep and continues without interruption to the front of the figurine, where the vulva with pronounced labia majora is visible between the open legs. There can be no doubt that the depiction of oversized breasts, accentuated buttocks and genitalia results from the deliberate exaggeration of the sexual features of the figurine.

In addition to the many carefully depicted anatomical features, the surface of the Venus preserves numerous lines and markings. The top of the Venus shows a series of U-shaped incisions on the roughly flat surface formed by the top of the breasts and the shoulders. The shoulders preserve multiple markings, with the short, deep, vertically incised lines along the back side of the figurine being the most pronounced. The breasts and arms also have multiple short, deeply incised lines that add to the three dimensionality of the sculpture. These markings are reminiscent of the various incisions found on other examples of ivory figurines from the Swabian Aurignacian, but, as is true of the others, this depiction is unique[2,15]. The Venus shows no signs of having been covered with pigments.

Many of the features, including the extreme emphasis on sexual attributes and lack of emphasis on the head, face and arms and legs, call to mind aspects of the Venus figurines well known from the European Gravettian, which typically date from between 22 and 27 kyr BP[16,17]. The careful depiction of the hands is reminiscent of those of Venuses such as the archetypal Venus of Willendorf—which was discovered 100 years earlier, in the summer of 1908—and a Venus from Kostenki I[17,18]. Despite the far greater age of the Venus of Hohle Fels, many of its attributes can be found in various forms in the rich tradition of Palaeolithic female representations. Although the Venus has numerous unique features, the presence of a ring for suspension in place of the head, and the upright, oversized breasts and massive shoulders relative to the flat stomach and small, pointed legs are particularly noteworthy.

The new figurine from Hohle Fels radically changes our view of the origins of Palaeolithic art. Before this discovery, animals and therianthropic imagery dominated the two dozen figurines from the Swabian Aurignacian. Female imagery was entirely unknown[2,15].

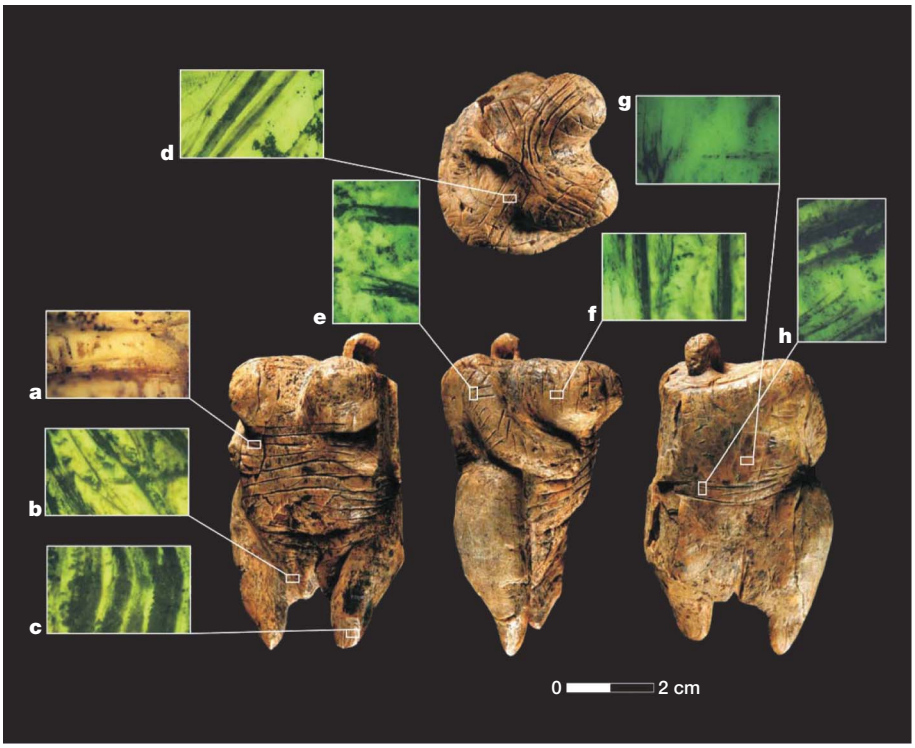


Figure 3 | Views of the Venus of Hohle Fels and photomicrographs documenting the methods of production. Multiple examples of cutting and incising (**a**–**f**, **h**) and surface polish (**g**). The photomicrographs were made with a Leica DMRX-MPV SP microscope photometer. **a**, Incident light, obliquely crossed polars, λ plate; **b**–**h**, Incident-light fluorescence mode (Ultraviolet- and violet-light excitation). In **a**–**d**, **f** and **g**, width of field of view is 2.7 mm; in **e** and **h**, width of field of view is 1.6 mm. Photographs by H. Jensen, photomicrographs by B. Ligouis; copyright, University of Tübingen.

With this discovery, the widespread notion that three-dimensional female depictions developed in the Gravettian can be rejected[17]. Interpretations suggesting that strong, aggressive animals or shamanic depictions dominate the Aurignacian art of Swabia, or even of Europe as a whole, must be reconsidered[15,19]. Although there is a long history of debate over the meaning of Palaeolithic Venuses, their clearly depicted sexual attributes suggest that they are a direct or indirect expression of fertility[20].

The stratigraphic position of the Venus of Hohle Fels indicates that it is the oldest of all of the figurines recovered from the Swabian caves and perhaps the earliest example of figurative art worldwide. The most noteworthy figurative representations of roughly comparable age outside Swabia are limited to the schematic, monochrome, red paintings on rock fragments from Fumane Cave in northern Italy[2,21], the standing figurine from Stratzing in the Wachau of Lower Austria[2,22] and the impressive paintings from Grotte Chauvet in the Ardèche in southern France[2,23]. Female imagery is rare in the early Upper Palaeolithic and includes a schematic example of parietal art from Chauvet, the figurine from Stratzing and engraved vulvas from several rock shelters in southwestern France[2,17,20,24]. The oldest evidence for figurative depictions outside Europe are seven paintings on mobile stone blocks from Apollo 11 Cave in southwestern Namibia, which date from between 25.5 and 27.5 kyr BP[25].

The Venus of Hohle Fels provides an entirely new view of the art from the early Upper Palaeolithic and reinforces the arguments that have been made for innovative cultural manifestations accompanying the rise of the Aurignacian in the upper Danube region[7,26]. Although the radiocarbon dates are ambiguous, as they often are in this period, the stratigraphic position of the figurine at the base of the thick Aurignacian deposits, which lack micro- and macroscopic signs of reworking, corroborate the abundant evidence for ivory working from the lower Aurignacian of Geißenklösterle and Hohle Fels. The archaeological context of the Venus of Hohle Fels indicates that innovations including the production of ivory figurines were present from the start of the Swabian Upper Palaeolithic[26]. Comparable depictions are entirely unknown at this early date, suggesting a local origin for this kind of female iconography.

No diagnostic human remains have been found in these strata[27,28]. Although I, as well as many other researchers, assume that the Aurignacian artworks were made by early modern humans shortly after their migration into Europe, this assumption can neither be confirmed nor refuted on the basis of the available skeletal data from the Swabian caves.

Received 24 January; accepted 17 March 2009.

1. Mellars, P., Boyle, K., Bar-Yosef, O. & Stringer, C. (eds) *Rethinking the Human Revolution* (McDonald Institute, 2007).
2. Floss, H. & Rouquerol, N. (eds) *Les Chemins de l'Art Aurignacien en Europe / Das Aurignacien und die Anfänge der Kunst in Europa* (Éditions Musée-forum Aurignac, 2007).
3. Bar-Yosef, O. & Zilhão, J. (eds) *Towards a Definition of the Aurignacian* (Proc. Symp. Trabalhos de Arqueologia 45, Instituto Português de Arqueologia/ American School of Prehistoric Research, 2006).
4. Henshilwood, C. S. *et al.* The emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* **295,** 1278–1280 (2002).
5. Parkington, J., Poggenpoel, C., Rigaud, J.-P. & Texier, P.-J. in *From Tools to Symbols: From Early Hominids to Modern Humans* (eds d'Errico, F. & Blackwell, L.) 475–492 (Witwatersrand University Press, 2005).
6. Brock, F., Bronk Ramsey, C. & Higham, T. Quality assurance of ultrafiltered bone dating. *Radiocarbon* **49,** 187–192 (2007).
7. Conard, N. J. & Bolus, M. Radiocarbon dating the appearance of modern humans and the timing of cultural innovations in Europe: new results and new challenges. *J. Hum. Evol.* **44,** 331–371 (2003).
8. Conard, N. J. & Bolus, M. Radiocarbon dating the late Middle Paleolithic and the Aurignacian of the Swabian Jura. *J. Hum. Evol.* **55,** 886–897 (2008).
9. Conard, N. J. Palaeolithic ivory sculptures from southwestern Germany and the origins of figurative art. *Nature* **426,** 830–832 (2003).
10. Richter, D., Waiblinger, J., Rink, W. J. & Wagner, G. A. Thermoluminescence, electron spin resonance and ^{14}C-dating of the late Middle Paleolithic and the early Upper Paleolithic site of Geißenklösterle in southern Germany. *J. Archaeol. Sci.* **27,** 71–89 (2000).
11. Zilhão, J. & d'Errico, F. in *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (eds Zilhão, J. & d'Errico, F.) 313–349 (Proc. Symp. Trabalhos de Arqueologia 33, Instituto Português de Arqueologia, 2003).

12. Jöris, O., &. Street, M. At the end of the ^{14}C time scale – the Middle to Upper Paleolithic record of western Eurasia. *J. Hum. Evol.* **55,** 782–802 (2008).
13. Weninger, B. & Jöris, O. A ^{14}C age calibration curve for the last 60 ka: the Greenland-Hulu U/Th timescale and its impact on understanding the Middle to Upper Paleolithic transition in Western Eurasia. *J. Hum. Evol.* **55,** 772–781 (2008).
14. Conard, N. J. & Bolus, M. in *Towards a Definition of the Aurignacian* (eds Bar-Yosef, O. & Zilhão, J.) 211–239 (Proc. Symp. Trabalhos de Arqueologia 45, Instituto Português de Arqueologia/American School of Prehistoric Research, 2006).
15. Hahn, J. *Kraft und Aggression. Die Botschaft der Eiszeitkunst im Aurignacien Süddeutschlands?* (Archaeologica Venatoria, 1986).
16. Roebroeks, W., Mussi, M., Svoboda, J. & Fennema, K. (eds) *Hunters of the Golden Age* (Analecta Praehistorica Leidensia, 2000).
17. Svoboda, J. in *Petřkovice: on Shouldered Points and Female Figurines* 193–223 (Institute of Archaeology at Brno, Academy of Sciences of the Czech Republic, 2008).
18. Antl-Weiser, W. *Die Frau von W. Die Venus von Willendorf, ihre Zeit und die Geschichte(en) um ihre Auffindung* (Verlag des Naturhistorischen Museums Wien, 2008).
19. Lewis-Williams, D. *The Mind in the Cave* (Thames and Hudson, 2002).
20. Delporte, H. *Image de la Femme dans l'Art Préhistorique* (Picard, 1993).
21. Broglio, A. Paleolithico e Mesolitico. *Preistoria Veronese* **5,** 11–56 (2002).
22. Neugebauer-Maresch, C. Zum Neufund einer weiblichen Statuette bei den Rettungsgrabungen an der Aurignacien-Station Stratzing/Krems-Rehberg, Niederösterreich. *Germania* **67,** 551–559 (1989).
23. Clottes, J. (ed.) *La Grotte Chauvet: L'Art des Origines* (Seuil, 2001).
24. Leroi-Gourhan, A. *Préhistoire de l'Art Occidental* Revised edn (Citadelles & Mazenod, 1995).
25. Vogelsang, R. *Middle-Stone-Age-Fundstellen in Südwest-Namibia* (Heinrich Barth Institut, 1998).
26. Conard, N. J., Dippon, G. & Goldberg, P. in *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (eds Zilhão, J. & d'Errico, F.) 165–176 (Proc. Symp. Trabalhos de Arqueologia 33, Instituto Português de Arqueologia, 2003).
27. Churchill, S. E. & Smith, F. H. Makers of the Early Aurignacian of Europe. *Yearb. Phys. Anthropol.* **43,** 61–115 (2001).
28. Conard, N. J., Grootes, P. M. & Smith, F. H. Unexpectedly recent dates for human remains from Vogelherd. *Nature* **430,** 198–201 (2004).

Acknowledgements Many colleagues, including S. Bailey, H. Bocherens, M. Bolus, S. Feine, H. Floss, P. Goldberg, P. Grootes, B. L. Hardy, T. Higham, M. Hofreiter, P. Krönneck, M. Kucera, L. Moreau, S. C. Münzel, D. Richter, F. H. Smith, H.-P. Uerpmann and S. Wolf have contributed to this research. I am particularly indebted to M. Malina for assistance during excavation and laboratory work, to R. Ehmann for the conservation of the Venus, to C. E. Miller for discussions on stratigraphy and to B. Ligouis for his microscopic images of the Venus. This research has been supported by the Deutsche Forschungsgemeinschaft, the University of Tübingen, the Heidelberger Akademie der Wissenschaften, the Landesamt für Denkmalpflege Baden-Württemberg, the Alb-Donau-Kreis, Heidelberg Cement, the Museumsgesellschaft Schelklingen and the Gesellschaft für Urgeschichte.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.J.C. (nicholas.conard@uni-tuebingen.de).

LETTERS

Snowdrift game dynamics and facultative cheating in yeast

Jeff Gore[1], Hyun Youk[1] & Alexander van Oudenaarden[1]

The origin of cooperation is a central challenge to our understanding of evolution[1–3]. The fact that microbial interactions can be manipulated in ways that animal interactions cannot has led to a growing interest in microbial models of cooperation[4–10] and competition[11,12]. For the budding yeast *Saccharomyces cerevisiae* to grow on sucrose, the disaccharide must first be hydrolysed by the enzyme invertase[13,14]. This hydrolysis reaction is performed outside the cytoplasm in the periplasmic space between the plasma membrane and the cell wall. Here we demonstrate that the vast majority (~99 per cent) of the monosaccharides created by sucrose hydrolysis diffuse away before they can be imported into the cell, serving to make invertase production and secretion a cooperative behaviour[15,16]. A mutant cheater strain that does not produce invertase is able to take advantage of and invade a population of wild-type cooperator cells. However, over a wide range of conditions, the wild-type cooperator can also invade a population of cheater cells. Therefore, we observe steady-state coexistence between the two strains in well-mixed culture resulting from the fact that rare strategies outperform common strategies—the defining features of what game theorists call the snowdrift game[17]. A model of the cooperative interaction incorporating nonlinear benefits explains the origin of this coexistence. We are able to alter the outcome of the competition by varying either the cost of cooperation or the glucose concentration in the media. Finally, we note that glucose repression of invertase expression in wild-type cells produces a strategy that is optimal for the snowdrift game—wild-type cells cooperate only when competing against cheater cells.

Yeast prefers to use the monosaccharides glucose and fructose as carbon sources. However, when these sugars are not available, yeast can metabolize alternative carbon sources such as the disaccharide sucrose[18]. After sucrose is hydrolysed by invertase, the resulting monosaccharides are imported[13,14], yet some of the glucose and fructose may diffuse away from the cell before it is able to import them into the cytoplasm (Supplementary Fig. 1). If such sugar loss by diffusion is significant then we might expect high-density cultures to grow more quickly than low-density cultures, because cells at high density benefit from their hydrolysis products and those of their abundant neighbours. Indeed, we find that cells grown in media supplemented with sucrose—but not glucose—grow much faster at high cell density than at low cell density. The growth rate at high cell density in 5% sucrose is similar to the growth rate at saturating (2%) glucose concentrations. However, the growth rate at low cell density is ~40% lower, equivalent to the growth rate in only 0.003% glucose (Supplementary Fig. 2). The fraction of invertase-created glucose that is captured can be estimated by dividing the rate of glucose uptake of cells growing in 0.003% glucose by the measured rate of invertase activity, yielding an estimated glucose capture efficiency of only ~1% (Supplementary Fig. 3). Analytic calculations of glucose diffusion suggest that this low capture efficiency is an expected consequence of diffusion and the known properties of the sugar importers (Supplementary Fig. 4).

Given that 99% of glucose created by a cell is lost to neighbouring cells, it may be possible for a 'cheater' strain to take advantage of the cooperators by not secreting invertase and instead simply consuming the glucose created by other cells[15]. If cooperative cells shared all of the glucose that they created (that is, if 100% of hydrolysed glucose and fructose diffused away from the hydrolysing cell), then both the cooperators and the cheaters would have the same access to sugar, yet only the cooperators would bear the metabolic cost of invertase production and secretion. In this case, the cheaters would always outgrow the cooperators, and the interaction would be what is called a prisoner's dilemma, in which cooperation is not sustainable in a well-mixed environment[1,17]. However, we found that yeast retains a small fraction of the glucose created by sucrose hydrolysis, which may be sufficient to allow cooperative strategies to survive.

To explore this problem, we performed a set of competition experiments between the wild-type strain ('cooperator') and a mutant strain lacking the invertase gene ('cheater' or 'defector'; see Supplementary Fig. 1). Consistent with there being a metabolic cost associated with invertase production, we find that in glucose-supplemented media, cooperators grow more slowly than cheaters only when invertase is being expressed (Supplementary Fig. 5)[15]. In addition, the cooperator strain in our experiments is a histidine auxotroph; therefore, limiting the histidine concentration in the media slows the growth of the cooperator relative to the cheater, allowing us to experimentally increase the 'cost of cooperation' (Supplementary Fig. 6). We can measure the relative abundance ('fractions') of the two strains in a mixed culture by flow cytometry because they express different fluorescent proteins (Supplementary Fig. 7).

We began by monitoring the change over time in the fractions of cooperators and cheaters co-cultured in sucrose media. Each co-culture started from a different initial fraction of cooperators, and each day we performed serial dilutions into fresh media and measured the cell density and relative abundance of the two strains. In cultures starting with a small fraction of cheaters, the cheaters increased in frequency, consistent with the cheaters 'taking advantage' of the cooperators (Fig. 1a). However, when the initial fraction of cooperators was low, we found that the frequency of cooperators increased, suggesting that in the steady state there will be coexistence between the two strains. Indeed, the equilibrium fraction is independent of the starting fraction but depends upon the histidine concentration (Fig. 1b; the equilibrium fraction in saturating histidine was $f \approx 0.3$). As the cost of cooperation increased, we observed a decrease in both the equilibrium fraction of cooperators and the mean growth rate of the culture at equilibrium (Fig. 1c). A large cost of cooperation therefore allows the cheaters to dominate the population but also results in a low growth rate of both strains. Coexistence was also observed in continuous culture, meaning that the 'seasonality'

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

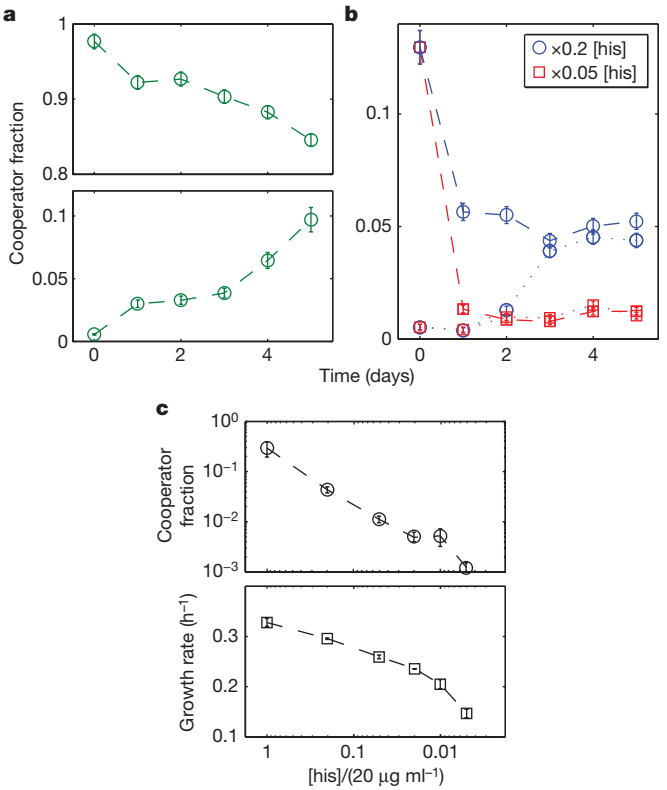


Figure 1 | Competition between the wild-type cooperator and mutant cheater strains. **a**, In sucrose culture, a small fraction of cheaters can invade a population of cooperators (top), and a small fraction of cooperators can also invade a population of cheaters (bottom), together implying coexistence between the two strains at steady state (histidine concentration ([his]), $20\,\mu g\,ml^{-1} \equiv \times 1$; no imposed cost of cooperation). **b**, As the histidine concentration becomes limiting we find that equilibrium between the two strains is reached within experimental timescales regardless of starting fractions. The fraction of cooperators at equilibrium does not depend upon the starting fraction but does depend upon the histidine concentration. **a** and **b** show typical data; error bars reflect sensitivity of measured fractions to different cut-off values (Supplementary Fig. 7). **c**, Both the equilibrium fraction of cooperators (circles) and the mean growth rate (squares) decrease as the cost of cooperation increases (lower histidine concentrations). Error bars, s.e.m.; $n = 3$.

imposed by serial dilution in batch culture is not necessary for coexistence[11] (Supplementary Fig. 8).

When the cooperators are initially only a small fraction of the population, then there will be little glucose available in the media. In this case, the cooperators have an advantage because they are able to capture at least some small fraction of the glucose that they create. As the fraction of cooperative cells increases, the glucose concentration also increases, and eventually the growth rates of the two strains become equal. Similarly, if the initial fraction of cooperative cells is above the equilibrium level, then their fraction will decrease; as this occurs, we find that the growth rate of the culture also decreases (Supplementary Fig. 9). Such a decrease in mean population fitness caused by evolutionary dynamics is a defining feature of the challenges posed by cooperation.

Our experimental observation of coexistence between the cooperator and cheater strains implies that the interaction is governed by what game theorists call the snowdrift game (also known as the hawk–dove game or the game of chicken)[3,17]. The snowdrift game derives its name from the potentially cooperative interaction present when two drivers are trapped behind a large pile of snow, and each driver must decide whether to clear a path. In this model of cooperation, the optimal strategy is the opposite of the opponent's (cooperate when your opponent defects and defect when your opponent cooperates). The snowdrift game is therefore qualitatively distinct from the prisoner's dilemma, in which all players have the incentive to cheat regardless of the strategies being followed by the others. Coexistence between cooperation and defection arises in a snowdrift game because rare strategies, which will often interact with the opposite strategy, do comparatively well.

To understand why sucrose metabolism is a snowdrift game, we constructed a simple phenomenological game theory model of the interaction. We assumed that invertase expression has a cost c and generates total benefits of unity that are captured with efficiency ε. In this scheme, for large capture efficiencies and/or small costs of cooperation ($\varepsilon > c$), the cooperators always outgrow the defectors and therefore take over the population (Fig. 2a). However, for small capture efficiencies and/or large costs ($\varepsilon < c$), the interaction is a prisoner's dilemma in which the defectors always do better, leading to extinction of the cooperators. However, in our experiments we observed coexistence between the two strains, an outcome that never occurs in the simple model of Fig. 2a. The ability to capture a sufficiently large fraction of the benefits of cooperation can allow cooperators to take over a population, but does not on its own lead to coexistence between cooperators and cheaters.

Coexistence of the two opposing strategies requires that the strains are mutually invasible. In particular, a lone cooperator must outperform a population composed entirely of defectors[17]. Indeed, we have already found experimentally that wild-type yeast in dilute cellular conditions is able to grow at a significant rate despite capturing only ~1% of the glucose created (Supplementary Fig. 2). This is because growth as a function of glucose is highly concave; doubling the glucose concentration therefore does not double the growth rate. By measuring the growth rate as a function of glucose concentration, we conclude that all benefit terms in our model should be raised to the power of $\alpha = 0.15 \pm 0.01$ (Supplementary Fig. 10 and Fig. 3c). Including this nonlinear effect alters the phase diagram and creates a large region of parameter space that is a snowdrift game in which there is coexistence between the two strategies[19] (Fig. 2b; $\alpha > 1$ leads to bistability[19,20] (Supplementary Table 1)). The saturating nature of growth on glucose means that a small number of cooperators can supply the glucose for many cells, thus providing a natural explanation for the small fraction of cooperators often observed in our competition experiments (Figs 1c and 3a, b).

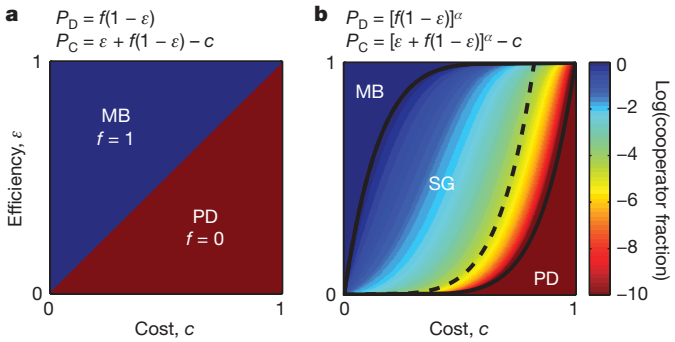


Figure 2 | Game theory models of cooperation in sucrose metabolism. **a**, Defection and cooperation payouts, respectively P_D and P_C, and the resulting phase diagram of the cooperative fraction, f, at equilibrium in a simple linear model in which cooperation has a cost c and leads to total benefits of unity that are captured with an efficiency ε. This model leads to fixation of cooperators ($f = 1$) at low cost and/or high efficiency of capture ($\varepsilon > c$, implying that the game is mutually beneficial (MB)[5]) but fixation of defectors ($f = 0$) for high cost and/or low efficiency of capture ($\varepsilon < c$, implying that the game is prisoner's dilemma (PD)). **b**, A model of cooperation with experimentally measured concave benefits yields a central region of parameter space that is a snowdrift game (SG), thus explaining the coexistence that is observed experimentally ($\alpha = 0.15$ in figure; see Supplementary Fig. 10). Adding glucose makes the cheaters less reliant on the cooperators, thus reducing the range of parameters in which cooperation can survive (solid to dashed line; see Supplementary Fig. 11).

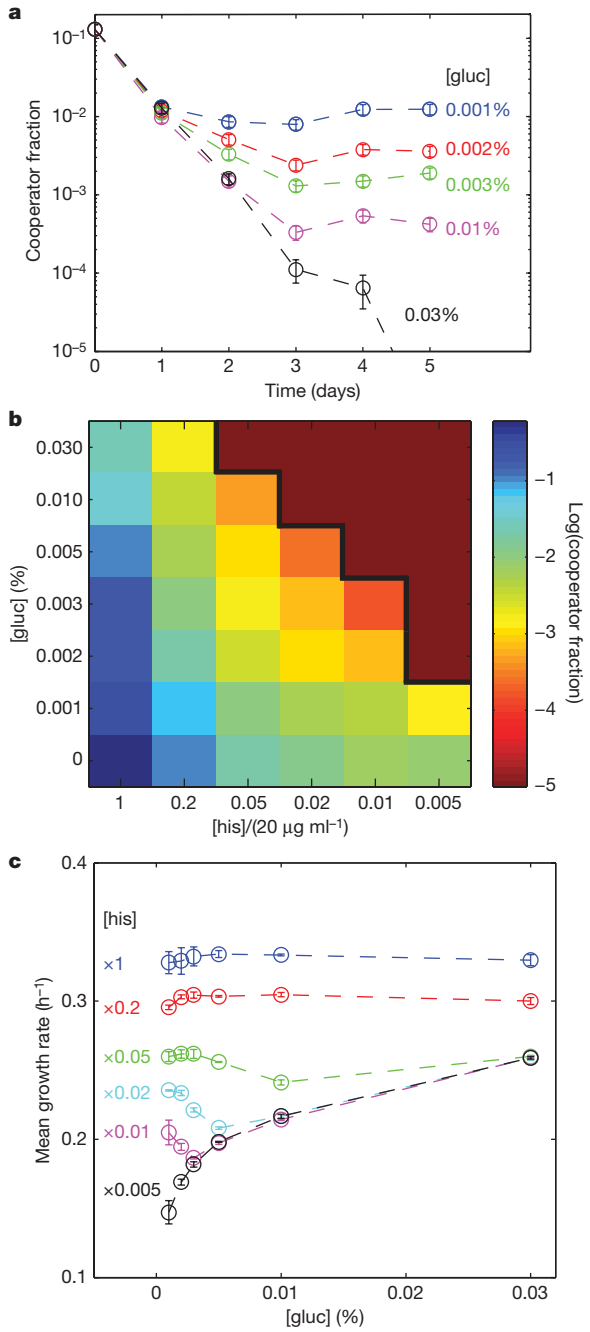


Figure 3 | Varying the glucose concentration can transform the outcome of competition. **a**, As the glucose concentration ([gluc]) in the media increases, the equilibrium fraction of cooperators decreases ([his] = ×0.05 = 1 μg ml^{-1}). Typical data shown; error bars reflect sensitivity of measured fractions to different cut-off values (Supplementary Fig. 7). **b**, Fraction of cooperators at equilibrium as a function of the glucose and histidine concentrations (all cultures have 5% sucrose; mean of two or three independent experiments; see Supplementary Fig. 12 for errors). The cooperators can be driven to extinction by either increasing the cost of cooperation or adding glucose to the media (solid black line denotes the extinction boundary). **c**, Mean growth rate of co-culture at equilibrium as a function of glucose concentration. Error bars, s.e.m.; $n = 3$. Adding glucose can decrease the growth rate at equilibrium because there are fewer cooperators to hydrolyse sucrose. As expected, if there are no cooperators at equilibrium then the growth rate is not a function of the histidine concentration. The nonlinear relationship between growth rate and glucose concentration is visible in the ×0.005 [his] data (black).

The sublinear relationship between growth rate and glucose suggests that the glucose concentration in the media may be an important parameter governing the cooperative interaction. As the glucose concentration increases, the cheaters become less reliant on the cooperators, and cooperation becomes more difficult to maintain (dashed line in Fig. 2b and Supplementary Fig. 11). Therefore, we expect that adding glucose will decrease the fraction of cooperators at equilibrium, eventually transforming the game into a prisoner's dilemma and driving the cooperators to extinction. The glucose concentration necessary to transform the game into a prisoner's dilemma is expected to be a decreasing function of the cost of cooperation. These predictions and the associated phase diagram can be confirmed experimentally (Fig. 3a, b and Supplementary Fig. 12).

Increasing the amount of glucose available in the media decreases the fraction of cooperators at equilibrium and can even drive the cooperators to extinction. As the cooperators decrease in frequency, the amount of sucrose being hydrolysed also decreases. We find that for some costs of cooperation, this effect is so severe that the equilibrium growth rate of the mixed culture actually decreases as we add glucose to the media (Fig. 3c). This non-intuitive decrease in the co-culture growth rate is a striking result of the cooperative interaction, as the growth rate of each strain cultured alone increases as glucose levels increase in the media.

Similar to many other alternative modes of carbon metabolism, invertase expression is repressed at high concentrations of glucose[18]. Given this genetically encoded strategy, we can ask how a wild-type cell responds when placed in competition against cells that either always cooperate or always defect. Competition against always-defecting cells leads to low glucose concentrations, resulting in wild-type cells cooperating by expressing invertase (as in our competition experiments). By contrast, a wild-type cell competing against an always-cooperating strain would result in the glucose concentration rising to the point (>0.1%) at which invertase expression is repressed, thus causing the wild-type cell to cheat[18,21] (Supplementary Fig. 5a). We therefore see that the wild-type invertase-production strategy is exactly what might be expected in a snowdrift game—wild-type cells pursue the strategy opposite to that of their opponents. It is possible that glucose repression of invertase is partly determined by these social considerations, helping to make a population of wild-type cells relatively immune to invasion by strains with alternative strategies[22].

Our results are consistent with a recent study which found that a cheater strain was more fit than the wild-type cooperator strain when growing at high density on a sucrose plate[15]. In that paper, sucrose metabolism was classified as a prisoner's dilemma, although the experimental results are also consistent with the cooperative interaction being a snowdrift game. Distinguishing between these two games requires observation of competition at low starting fraction of cooperator. In addition, the competition must be performed in a well-mixed environment because spatial structure, such as the agar plate used in ref. 15, can drastically affect the outcome of competition[16,23].

The experimental observation of coexistence between cooperator and cheater strains in a well-mixed environment makes sucrose metabolism in yeast a particularly clear example of the snowdrift game[24], and may explain the existence in wild yeast populations of copy number variation in the *SUC2* gene, including the presence of cheaters[25]. Coexistence between cooperator and cheater strains in our experiments provide a concrete example of how interactions between alternative alleles can promote biological diversity[11,24,26]. Similar cooperative interactions may be present in other enzymatic processes that occur in the periplasmic space of yeast such as phosphate scavenging, starch degradation and phospholipase activity. It would be interesting to study the outcome of competition between the cooperator and cheater strains in spatially structured environments[9,15,16,23,27–29], particularly given a recent theoretical prediction that spatial structure often inhibits cooperation in a snowdrift game[27].

METHODS SUMMARY

Strains. All strains were derived from haploid cells BY4741 (mating type **a**, EUROSCARF). The 'wild-type' cooperator strain has an intact *SUC2* gene, a

defective *HIS3* gene (*his3Δ1*) and yellow fluorescent protein expressed constitutively by the *ADH1* promoter (inserted using plasmid pRS401 containing *MET17*). The mutant cheater strain lacks the *SUC2* gene (EUROSCARF Y02321, *SUC2::kanMX4*), has an intact *HIS3* gene, and has the fluorescent protein tdTomato expressed constitutively by the *PGK1* promoter (inserted using plasmid pRS301 containing *HIS3*). Growth rate and invertase expression experiments in Supplementary Figs 2 and 5a were done using a strain containing yellow fluorescent protein driven by the *SUC2* promoter (inserted using plasmid pRS306 containing *URA3*).

Competition experiments. Co-culture experiments were performed in 5 ml batch culture at 30 °C using synthetic media (minus histidine) supplemented with 5% sucrose and variable concentrations of glucose and histidine. Cultures were maintained in a 'well-mixed' condition by growing in an incubator shaker at 225 r.p.m. The 20% sucrose stock solution was filter-sterilized and stored with 1 mM Tris buffer, pH 8.0, to prevent acid-catalysed autohydrolysis. Nevertheless, 5% sucrose media typically had a monosaccharide concentration of ~0.0001%. The experiments described in Fig. 1 have 0.001% glucose added manually. Serial dilutions were performed daily (23 h of growth) such that the starting optical density was 0.0025, corresponding to ~150,000 cells. Fractions were determined using a BD FACScan flow cytometer (Supplementary Fig. 7) and periodically confirmed by selective plating. Equilibrium data in Figs 1c and 3b, c were recorded after five days of competition between the two strains.

Received 23 November 2008; accepted 11 February 2009.
Published online 6 April 2009.

1. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211,** 1390–1396 (1981).
2. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314,** 1560–1563 (2006).
3. Smith, J. M. *Evolution and the Theory of Games* 167–173 (Cambridge Univ. Press, 1982).
4. Velicer, G. J. Social strife in the microbial world. *Trends Microbiol.* **11,** 330–337 (2003).
5. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4,** 597–607 (2006).
6. Griffin, A. S., West, S. A. & Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature* **430,** 1024–1027 (2004).
7. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104,** 1877–1882 (2007).
8. Diggle, S. P., Griffin, A. S., Campbell, G. S. & West, S. A. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450,** 411–414 (2007).
9. Rainey, P. B. & Rainey, K. Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425,** 72–74 (2003).
10. Smukalla, S. *et al.* FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* **135,** 726–737 (2008).
11. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature* **441,** 498–501 (2006).
12. Chao, L. & Levin, B. R. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl Acad. Sci. USA* **78,** 6324–6328 (1981).
13. Carlson, M. & Botstein, D. 2 differentially regulated messenger-RNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. *Cell* **28,** 145–154 (1982).
14. Dickinson, J. R. & Schweizer, M. *The Metabolism and Molecular Physiology of* Saccharomyces cerevisiae 54–55 (CRC, 2004).
15. Greig, D. & Travisano, M. The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proc. R. Soc. Lond. B* **271** (suppl.), 25–26 (2004).
16. Maclean, R. C. & Brandon, C. Stable public goods cooperation and dynamic social interactions in yeast. *J. Evol. Biol.* **21,** 1836–1843 (2008).
17. Doebeli, M. & Hauert, C. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* **8,** 748–766 (2005).
18. Gancedo, J. M. Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* **62,** 334–361 (1998).
19. Motro, U. Co-operation and defection: playing the field and the ESS. *J. Theor. Biol.* **151,** 145–154 (1991).
20. Skyrms, B. *The Stag Hunt and Evolution of Social Structure* (Cambridge Univ. Press, 2004).
21. Ozcan, S., Vallier, L. G., Flick, J. S., Carlson, M. & Johnston, M. Expression of the SUC2 gene of *Saccharomyces cerevisiae* is induced by low levels of glucose. *Yeast* **13,** 127–137 (1997).
22. Santorelli, L. A. *et al.* Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature* **451,** 1107–1110 (2008).
23. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359,** 826–829 (1992).
24. Fiegna, F. & Velicer, G. J. Competitive fates of bacterial social parasites: persistence and self-induced extinction of Myxococcus xanthus cheaters. *Proc. R. Soc. Lond. B* **270,** 1527–1534 (2003).
25. Naumov, G. I., Naumova, E. S., Sancho, E. D. & Korhola, M. P. Polymeric SUC genes in natural populations of *Saccharomyces cerevisiae. FEMS Microbiol. Lett.* **135,** 31–35 (1996).
26. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418,** 171–174 (2002).
27. Hauert, C. & Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428,** 643–646 (2004).
28. Keymer, J. E., Galajda, P., Muldoon, C., Park, S. & Austin, R. H. Bacterial metapopulations in nanofabricated landscapes. *Proc. Natl Acad. Sci. USA* **103,** 17290–17295 (2006).
29. Nadell, C. D., Xavier, J. B. & Foster, K. R. The sociobiology of biofilms. *FEMS Microbiol. Rev.* **33,** 206–224 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank D. Kim, A. Raj, K. Gora, D. Muzzey and B. Pando for discussions and/or experimental help. This work was supported by grants from the US National Institutes of Health (NIH) and National Science Foundation to A.v.O. J.G. is supported through a Pappalardo Postdoctoral Fellowship and an NIH K99 Pathways to Independence Award. H.Y. was supported by a Lester Wolfe Fellowship.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.v.O. (avano@mit.edu).

LETTERS

Two-year-olds with autism orient to non-social contingencies rather than biological motion

Ami Klin[1], David J. Lin[1]†, Phillip Gorrindo[1]†, Gordon Ramsay[1,2] & Warren Jones[1,3]

Typically developing human infants preferentially attend to biological motion within the first days of life[1]. This ability is highly conserved across species[2,3] and is believed to be critical for filial attachment and for detection of predators[4]. The neural underpinnings of biological motion perception are overlapping with brain regions involved in perception of basic social signals such as facial expression and gaze direction[5], and preferential attention to biological motion is seen as a precursor to the capacity for attributing intentions to others[6]. However, in a serendipitous observation[7], we recently found that an infant with autism failed to recognize point-light displays of biological motion, but was instead highly sensitive to the presence of a non-social, physical contingency that occurred within the stimuli by chance. This observation raised the possibility that perception of biological motion may be altered in children with autism from a very early age, with cascading consequences for both social development and the lifelong impairments in social interaction that are a hallmark of autism spectrum disorders[8]. Here we show that two-year-olds with autism fail to orient towards point-light displays of biological motion, and their viewing behaviour when watching these point-light displays can be explained instead as a response to non-social, physical contingencies—physical contingencies that are disregarded by control children. This observation has far-reaching implications for understanding the altered neurodevelopmental trajectory of brain specialization in autism[9].

Preferential attention to biological motion is a fundamental mechanism facilitating adaptive interaction with other living beings. It is present throughout a wide range of species, from humans[10,11] to monkeys[12] to birds[13]. Developmentally, it can be found in newly hatched chicks[14] and in human infants as young as 2 days old[1]. Recognition of biological motion remains intact in a variety of forms, from degraded presentations, through varying states of occlusion, and in cases when information-bearing components are reduced to their most minimal[15,16]. In addition, perception of biological motion can be preserved even when other types of motion perception are impaired, as in individuals with Williams syndrome[17] (a condition noted for visuo-spatial deficits) and in patients suffering from circumscribed brain lesions[18]. Furthermore, biological motion perceived through other sensory modalities—such as when listening to sounds of human motion[19]—evokes activity in the same areas of the brain that are typically responsive to visual presentations.

Collectively, these findings describe a mechanism that is evolutionarily well-conserved, developmentally early-emerging, highly robust in signal detection (withstanding degradation on signalling and receiving sides), and redundantly represented by several sensory modalities. Each of these aspects suggests ready benefits for adaptive interaction with other living beings: following the movements of a conspecific, looking at others to entreat or avoid interaction, learning by imitation, or directing preferential attention to cues that build on biological motion (such as facial expression and gaze direction[5]).

Notably, many of the same behaviours have also been shown as deficits in children with autism: deficits in social interaction, diminished eye contact and reduced looking at others, problems with imitation, deficits in recognizing facial expressions, and difficulties following another's gaze[20]. Autism is a lifelong, highly prevalent, and strongly genetic disorder defined by impairments in social and communicative functioning and by pronounced behavioural rigidities[21]. Although the preponderance of evidence points to prenatal factors instantiated in infancy, knowledge of the first two years of life in autism remains largely limited to retrospective data and indirect observations[20]: because autism is rarely diagnosed before 18 months, relatively little is known about autism during the first two years of development.

In later life, much more is known about the consequences—cognitive, social and behavioural—of having autism. Altered visual scanning, of both faces and social scenes[22,23], as well as altered neural processing of social information, have been documented[24,25]. In school-age children with autism, perception of biological motion is impaired[26], but the manner in which very young children with autism relate to biological motion in early life, during periods critical for brain development and before compensatory coping strategies are established, has not, to our knowledge, been previously studied.

In the current study, we sought to address whether preferential attention to biological motion is altered in children with autism by two years of age, and what other factors might guide the visual attention of children with autism if they do fail to orient towards biological motion.

To answer these questions, we created five sets of point-light animations, counterbalanced for a total of ten. The animations consisted of children's games, such as playing 'peek-a-boo' or 'pat-a-cake', and were created with live actors and motion capture technology (see Supplementary Information). The motion capture sessions included a simultaneous audio recording. The experimental task was a preferential-looking paradigm (Fig. 1a and Supplementary Movie 1): a point-light animation of biological motion was presented on one half of a computer screen, together with the audio soundtrack of the actor's vocalizations. On the other half of the screen, the same animation was presented, but that point-light figure was inverted in orientation (shown upside-down) and played in reverse order (the frames of animated action played from the end of the sequence until its beginning). Only the one (forward) audio soundtrack was presented.

Inverted presentation disrupts perception of biological motion in young children[27], and is processed by different neural circuits in infants as young as 8 months old[28]. Also, by playing the inverted

[1]Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut 06519-1124, USA. [2]Haskins Laboratories, New Haven, Connecticut 06511, USA. [3]Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, Connecticut 06520-8074, USA. †Present addresses: Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA (D.J.L.); Neuroscience Graduate Program at Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, Tennessee 37203, USA (P.G.).

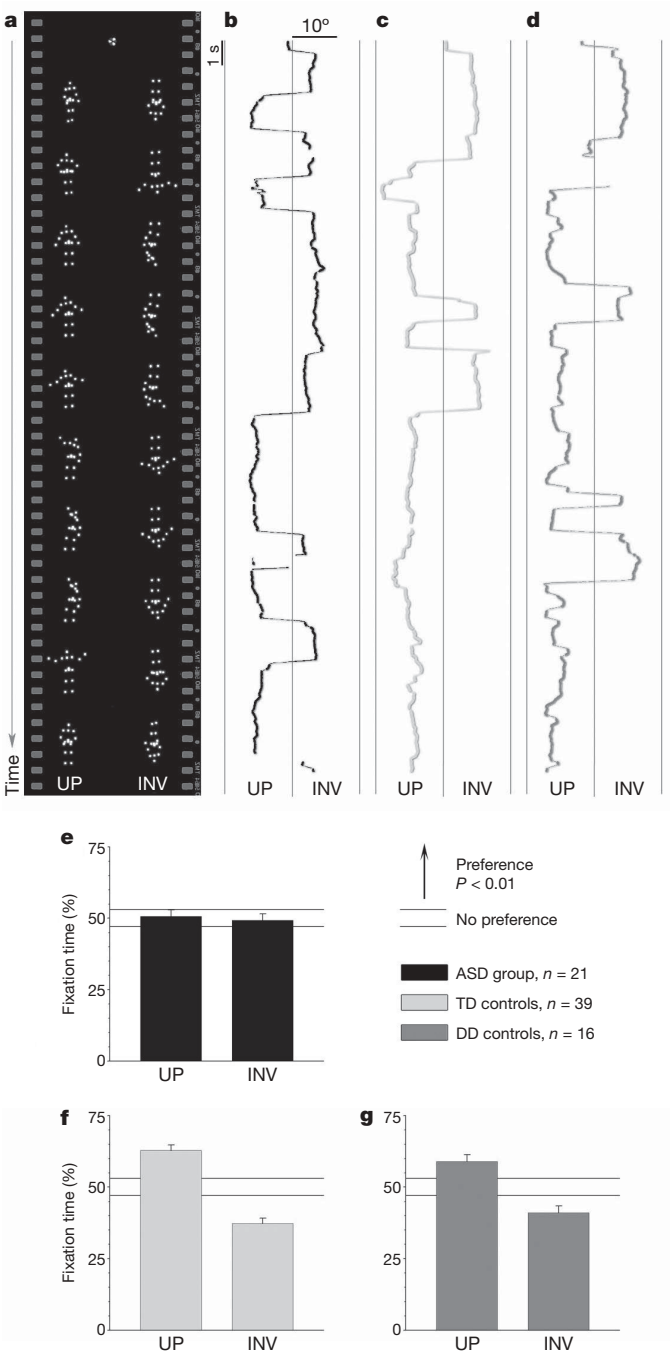


Figure 1 | Two-year-olds with autism show no preferential attention to biological motion, whereas control children show significant preferences. **a**, Example still images from point-light biological motion stimuli, with centring cue at start. Each animation showed an upright (UP) and inverted (INV) figure with accompanying soundtrack matching the actions of the upright figure. The upright figure enacted childhood games. Figures were identical except that the inverted figure was rotated 180° and its movements were played in reverse order. **b**–**d**, Visual scanning data of individual children are plotted as horizontal location by time. Breaks in the data occur for blinks or offscreen fixations. **b**, Visual scanning data from one toddler with autism (ASD), for one animation. **c**, Data from one typically developing toddler (TD). **d**, Data from one developmentally delayed but non-autistic toddler (DD). **e**, For the ASD group, fixation to upright and inverted biological motion occurs at chance levels. **f**, Typically developing toddlers give preferential attention to upright animations. **g**, Developmentally delayed toddlers also give preferential attention to upright animations. Horizontal guidelines denote percentages not significantly different from chance. Error bars are s.e.m.

animation backwards, its relative levels of motion complexity, speed and gestalt coherence were preserved, but its motion was not an exact mirror of the upright. Each animation lasted an average of 30 s. The order of presentation was randomized, and the presentation of the upright figure was counterbalanced to appear on the left and right side of the screen equally often.

Evidence for recognition and preferential attention to biological motion was measured by the child's viewing patterns: increased looking towards the upright figure indicated preferential attention to biological motion[1] and the perceptual matching of human voice with a mental template of human action[8]. Visual scanning was measured with eye-tracking equipment, with data collected at 60 Hz (Fig. 1b–d) (see Methods in Supplementary Information).

With the written, informed consent of their parents or legal guardians, 76 children with a mean chronological age of 2.05 (s.d. = 0.62) participated. These children comprised three groups (see Supplementary Table 1): 21 toddlers with autism spectrum disorders (ASD), 39 typically developing toddlers, and 16 developmentally delayed but non-autistic toddlers. Toddlers with autism were matched to the typically developing toddlers on non-verbal mental age and chronological age, and matched to the developmentally delayed, non-autistic toddlers on verbal mental age and chronological age (see Supplementary Information).

Whereas typically developing toddlers provide normative data, the developmentally delayed but non-autistic children act as controls against developmental confounds, assuring that the findings are specific to autism rather than attributable to delays in cognitive development or language function.

Results are plotted in Fig. 1e–g. When viewing point-light displays of human biological motion, two-year-olds with autism spectrum disorders are random in their looking patterns: 50.7% on the upright figure versus 49.3% on the inverted (Fig. 1e). In contrast, both control groups demonstrated significant preferential attention to the upright animations: 62.7% upright for the typically developing group, and 58.9% upright for the developmentally delayed group (Fig. 1f, g). Comparison across groups was significantly different (by one-way analysis of variance (ANOVA), $F_{2,73} = 7.95$, $P < 0.001$). In pairwise comparisons, looking by the ASD group differed significantly from that of each control group ($P < 0.001$ in comparison with the typically developing group, and $P = 0.0185$ relative to the developmentally delayed group). The two control groups did not differ significantly from one another ($P = 0.27$). All data were normally distributed (all $P > 0.4$, $k < 0.15$, Lilliefors).

Results in Fig. 1 are from four of the five types of animations presented. In earlier research[7], a serendipitous observation led us to recognize that one of the animations contained a confounder. Although four animations presented only moving point-lights with an accompanying human voice, one animation included a different sound. The actor in that animation plays pat-a-cake (see Supplementary Movie 2), and the sound of clapping is heard at the same time that two point-lights—the actor's hands—collide. The collision of point-lights and the resulting clapping sound create a causal physical contingency: rather than merely co-occurring (as with the speech sounds and movements in the other animations), the movements of the point-light hands in this case actually cause a noise to occur. In the earlier research we found that a 15-month-old with autism was very sensitive to the occurrence of this clapping, as her preferential looking went from random during other animations to 93.1% upright during the pat-a-cake animation[7].

During the clapping, the causal physical contingency only exists on the upright side: the single audio track plays normally (forward), matching the upright movements, but the action of the inverted figure, playing in reverse, does not move in time to the clapping sounds.

When analysed independently (Fig. 2), the toddlers with ASD showed a significant preference for the upright clapping figure during the pat-a-cake animation, and looking towards the upright figure during this animation was significantly increased relative to other

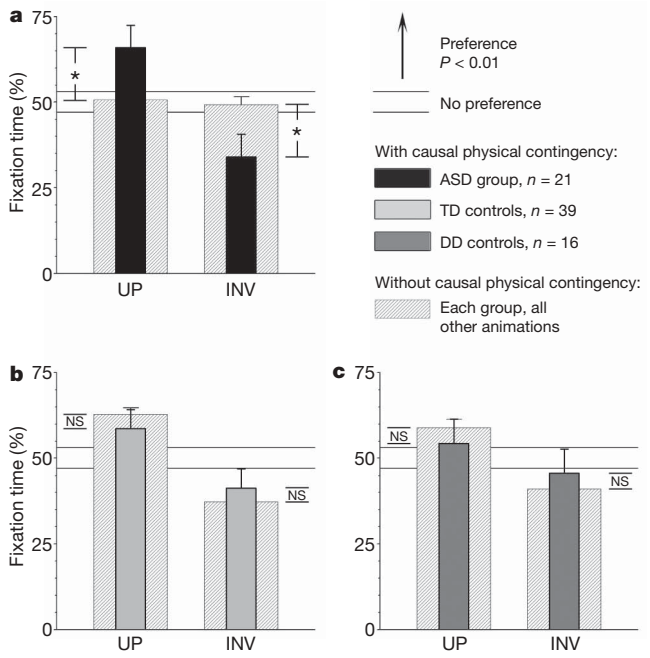


Figure 2 | When the animation contains a physical contingency, two-year-olds with autism do show significant viewing preferences. **a**, During other biological motion animations, ASD toddlers show no preference, but when a physical contingency is present on the upright side, these toddlers show significant preference for the upright figure (different from chance: $P < 0.01$; different from their viewing behaviour to other animations: $P = 0.044$). Whereas other animations presented only moving point-lights and human voice, one type of animation contained an extra cue: as two point-lights, representing the actor's hands, collided, the sound of clapping could be heard (playing 'pat-a-cake'). The collision of point-light 'hands' actually caused a noise (the clap) to occur, localized to the upright (UP) figure and absent from the inverted (INV; the inverted figure's movements were not synchronous with the claps). **b**, Typically developing toddlers show no significant change in preferential viewing. **c**, Developmentally delayed toddlers also show no significant change in preferential viewing. Horizontal guidelines denote percentages not significantly different from chance. Error bars are s.e.m. $*P < 0.05$. See Supplementary Movies for movie data.

animations: 65.9% upright during pat-a-cake versus only 50.7% in the other four animations, $t_{20} = 2.43$, $P = 0.02$. Behaviour of the typically developing and developmentally delayed groups was unchanged: they continued to give preferential attention to the upright figure: 58.6% upright during pat-a-cake versus 62.7% in the other four animations for typically developing ($t_{38} = 0.79$, $P = 0.44$); and 54.4% versus 58.9% for developmentally delayed ($t_{15} = 0.66$, $P = 0.51$). Overall on this animation, results for the three groups did not differ significantly ($F_{2,73} = 0.67$, $P = 0.52$). All data were normally distributed (all $P > 0.36$, $k < 0.15$, Lilliefors).

After this observation, we questioned whether the presence of more subtle synchronies might have had an unanticipated role in the viewing of all animations—that is, whether visual scanning that had appeared random by the toddlers with ASD might actually be related to audiovisual synchronies less obvious than clapping.

To test this, we quantified levels of audiovisual synchrony (AVS) in all animations (Fig. 3). In the pat-a-cake animation, when the point-light hands collide and a clapping sound occurs, an abrupt change in motion coincides with a large change in sound amplitude. We measured AVS in our stimuli to match this case: the synchronous occurrence of change in motion and change in sound[29].

We measured the change in motion by first measuring each point-light's trajectory over time (Fig. 3a). From each point-light's trajectory, we calculated its velocity and then the magnitude of its change in velocity, $|\Delta v|$ (Fig. 3b, c). This served as our measure of change in motion. To measure change in sound, we measured the audio amplitude of the

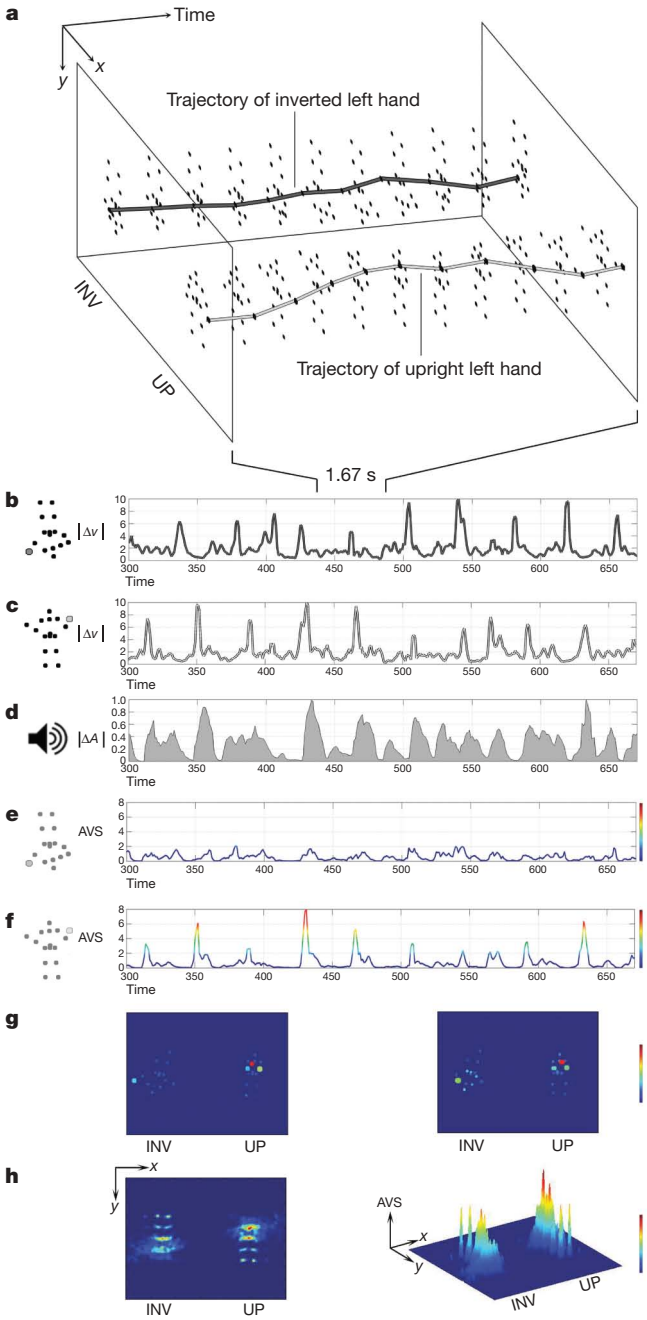


Figure 3 | Quantification of audiovisual synchrony. **a**, We measured spatial trajectories (x-y location over time) of all point-lights throughout each biological motion animation. Example trajectories are for inverted (INV) left hand and for upright (UP) left hand. **b**, Magnitude of change in velocity of inverted left hand, $|\Delta v|$. **c**, Magnitude of change in velocity of upright left hand. **d**, Magnitude of change in short-term amplitude envelope of audio soundtrack, $|\Delta A|$. **e**, AVS of inverted left hand, obtained as a pointwise product of **b** and **d**. **f**, AVS of upright left hand, obtained as a pointwise product of **c** and **d**. **g**, Two still frames from pat-a-cake animation. Colour scale values range from low or no synchrony (dark blue) to maximum synchrony (red). Note that some point-lights are very synchronous (the hands, shown here during claps), whereas others are hardly synchronous (for example, the feet). **h**, Summation of AVS over the duration of an entire animation. Oblique view shows that although there is more AVS on the upright side, the inverted side also contains synchrony: by chance alignment (reverse motion signal aligned with forward audio signal), some change in movement of inverted point-lights can occur synchronously with the change in audio. If preferential viewing in our stimuli were related to the level of AVS, then the relative levels of synchrony on the upright versus inverted side will provide predictions of expected viewing behaviour.

soundtrack (its short-term amplitude envelope) and then calculated its rate of change, the magnitude of ΔA, $|\Delta A|$ (Fig. 3d). The level of AVS of each point-light was then calculated as the product of change in velocity and change in sound amplitude (Fig. 3e, f). This measure of AVS was computed for all point-lights on both the upright and inverted sides (see Supplementary Movie 3 and Supplementary Information).

By then summing the AVS signals of all point-lights over time, we generated cumulative maps of AVS for each animation (Fig. 3h). From these maps, we calculated the difference between maximum AVS on the upright side and maximum synchrony on the inverted side (as a percentage difference to normalize across animations).

Across different animations, this measure of upright versus inverted synchrony then acted as a prediction of which side of the animation would be preferentially attended—if the viewing patterns of children were related to attention to AVS. The relationship between synchrony and preferential viewing was tested by regression (Fig. 4). For the ASD group, preferential looking was significantly and strongly correlated with level of AVS ($R^2 = 0.90$ and $P = 0.01$; Fig. 4a). In the typically developing and the developmentally delayed groups, there was no significant correlation between viewing and AVS ($R^2 = 0.29$ and 0.17, respectively; Fig. 4b, c). Correlation coefficients for the three groups were significantly different from one another ($\chi^2 = 7.24$, $P < 0.05$)[30], with the r value of the ASD group differing from that of the typically developing group ($z = 2.41$, $P < 0.05$) as well as the developmentally delayed group ($z = -2.25$, $P < 0.05$). The two control groups did not differ significantly. The pat-a-cake animation had the greatest upright AVS. When we withheld that animation and re-analysed, the correlation between preferential viewing and AVS remained significant for the ASD group ($R^2 = 0.95$ and $P = 0.018$), but was still not significant for the other groups ($R^2 = 0.04$ for typically developing, and $R^2 = 0.001$ for developmentally delayed).

The results from this *post hoc* quantification of AVS and preferential viewing indicated that the viewing patterns of toddlers with autism—random relative to social content—showed instead a marked reliance on AVS. This one measure accounted for 90% of the autism group's variance in preferential viewing. In contrast, the looking patterns of typically developing and of developmentally delayed, non-autistic children showed no relationship with the levels of AVS. The control children gave preferential attention to biological motion, disregarding AVS in favour of more socially relevant signals.

To test whether AVS could predict looking behaviour in new animations, we designed a follow-up experiment (Fig. 4d) in which we created two new types of animations with increased AVS levels, filling the gap in synchrony signal strength of our original stimuli. We recruited ten additional toddlers with ASD, characterized in the same manner and matched to the original ASD cohort (see Supplementary Information). We used our original results to build a predictive model for expected behaviour, creating weighted binomial prediction intervals around the original regression line[31], with specific predictions for each animation. The probability of both results falling within their respective prediction intervals is equal to the probability of obtaining a value in one interval multiplied by the probability of obtaining a value within the other ($P = [0.1674 - 0.0002] \times [0.0024 - 0]$).

Preferential viewing by this second cohort of toddlers with autism, watching new animations, fit the predictions on the basis of AVS: their viewing on each animation followed the model, a result with a chance likelihood of $P = 0.0004$.

Overall, these results indicate that a skill present in two-day-old, typically developing infants[1], as well as in chronologically, non-verbally, and verbally matched control children (the typically developing and developmentally delayed groups herein), is not functioning properly in children with autism at the age of two.

There are likely to be significant implications of a disruption to such a basic and highly conserved mechanism. One immediate implication of this finding concerns our understanding of another very basic behaviour: how infants with autism look at the faces of other people. We recently found that in comparison with control children, two-year-olds with autism look less at the eyes of others and attend instead to their mouths[24]. The present results indicate an explanation: where on the face is there greatest AVS? These children's sensitivity to synchrony in the present biological motion stimuli is consistent with fixating on the ongoing synchronies between lip motion and speech sound, and the lack of preferential attention towards biological motion is consistent with diminished attention to the eyes and diminished expertise in social action and interaction found in later life.

Developmentally, these results mark an important, early point along an alternative path of neural and behavioural specialization. Although individual and species-specific genetics begin the development of mind and brain, that development over time is shaped critically by experience. For infants with autism, this would suggest that genetic predispositions are probably exacerbated by experiences that are increasingly atypical. By two-years-of-age, the data in this report show that these children are on a substantially different developmental course, having learned already from a world in which the physical contingencies of coincident light and sound are quantifiably more salient than the rich social information imparted by biological motion. Future investigations will benefit from studies, starting still earlier in life, of the developmental unfolding of such selective learning profiles. Exactly which signals are spontaneously attended to and which are missed, and the consequences thereof for structural and

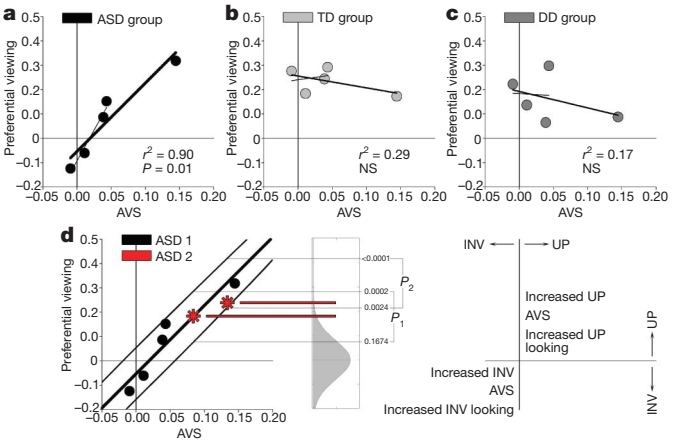


Figure 4 | The level of AVS is highly correlated with preferential viewing in two-year-olds with autism, is uncorrelated with viewing in control children, and can predict ASD viewing patterns in new animations. **a**, Preferential viewing is significantly correlated with AVS in ASD toddlers. Plots pair preferential viewing and AVS. When the animation with greatest upright AVS (pat-a-cake) is withheld from analysis, AVS is still significantly correlated with viewing behaviour in ASD toddlers: $r^2 = 0.95$, $P = 0.018$ (plotted as thin regression line through remaining four data points). **b**, Preferential viewing by typically developing (TD) toddlers is uncorrelated with AVS, across either four or five animations. **c**, Preferential viewing by developmentally delayed (DD) toddlers is also uncorrelated with AVS. **d**, To test whether AVS could predict looking behaviour in new animations, we created two further animation types. The regression from the original data, with weighted binomial prediction intervals, provided a model for expected behaviour. P_1 and P_2 denote prediction intervals for the new animations. Probability of obtaining the results in these intervals is noted to the right of the regression plot. For an independent cohort of toddlers with autism, matched to the original cohort, preferential viewing was predicted on the basis of AVS ($P = 0.0004$). In all plots, the y-axis shows preferential viewing as a difference score: percentage of fixation time to upright (UP) minus percentage of fixation time to inverted (INV). Positive values indicate increased looking at the upright. Similarly, the x-axis shows AVS as synchrony of the upright (as percentage of total synchrony) minus synchrony of the inverted (also as percentage of total). Positive values indicate greater synchrony in the upright figure.

functional brain development, may shed light on the neurobiological anomalies that predispose these altered avenues of learning.

METHODS SUMMARY

Children were recruited through a federally funded STAART Center (Studies to Advance Autism Research and Treatment, NIMH U54-MH66494) based in the Autism Program of the Yale Child Study Center, New Haven. The research protocol was approved by the Human Investigations Committee of the Yale University School of Medicine, and families were free to withdraw from the study at any time. The children were shown counterbalanced presentations of each of five point-light biological motion animations (for a total of ten presentations in the original experiment), and two extra animations (four presentations in the follow-up experiment) (see Fig. 1a and Supplementary Movies 1–3). Preferential viewing in our design was a binary choice, upright versus inverted. Visual scanning was measured with eye-tracking equipment (ISCAN, Inc.). The equipment uses a dark pupil/corneal reflection technique with data collected at the rate of 60 Hz. Analysis of eye movements and coding of preferential fixation data were performed with software written in MATLAB.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 November 2008; accepted 3 February 2009.
Published online 29 March 2009; corrected 14 May 2009 (see full-text HTML version for details).

1. Simion, F., Regolin, L. & Bulf, H. A predisposition for biological motion in the newborn baby. *Proc. Natl Acad. Sci. USA* **105,** 809–813 (2008).
2. Regolin, L., Tommasi, L. & Vallortigara, G. Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Anim. Cogn.* **3,** 53–60 (2000).
3. Blake, R. Cats perceive biological motion. *Psychol. Sci.* **4,** 54–57 (1993).
4. Johnson, M. H. Biological motion: a perceptual life detector? *Curr. Biol.* **16,** R376–R377 (2006).
5. Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T. & McCarthy, G. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb. Cortex* **15,** 1866–1876 (2005).
6. Frith, C. D. & Frith, U. Interacting minds: a biological basis. *Science* **286,** 1692–1695 (1999).
7. Klin, A. & Jones, W. Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism. *Dev. Sci.* **11,** 40–46 (2008).
8. Klin, A., Jones, W., Schultz, R. T. & Volkmar, F. The enactive mind – from actions to cognition: lessons from autism. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **358,** 345–360 (2003).
9. Johnson, M. H. Functional brain development in humans. *Nature Rev. Neurosci.* **2,** 475–483 (2001).
10. Johansson, G. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14,** 201–211 (1973).
11. Fox, R. & McDaniel, C. The perception of biological motion by human infants. *Science* **218,** 486–487 (1982).
12. Oram, M. W. & Perrett, D. I. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* **76,** 109–129 (1996).
13. Omori, E. & Watanabe, S. Discrimination of Johansson's stimuli in pigeons. *Int. J. Comp. Psychol.* **9,** 92 (1996).
14. Vallortigara, G., Regolin, L. & Marconato, F. Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol.* **3,** e208 (2005).
15. Thompson, J. C., Clarke, M., Stewart, T. & Puce, A. Configural processing of biological motion in human superior temporal sulcus. *J. Neurosci.* **25,** 9059–9066 (2005).
16. Neri, P., Morrone, M. C. & Burr, D. C. Seeing biological motion. *Nature* **395,** 894–896 (1998).
17. Jordan, H., Reiss, J. E., Hoffman, J. E. & Landau, B. Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychol. Sci.* **13,** 162–167 (2002).
18. Jokisch, D., Troje, N. F., Koch, B., Schwarz, M. & Daum, I. Differential involvement of the cerebellum in biological and coherent motion perception. *Eur. J. Neurosci.* **21,** 3439–3446 (2005).
19. Bidet-Caulet, A., Voisin, J., Bertrand, O. & Folumpt, P. Listening to a walking human activates the temporal biological motion area. *Neuroimage* **28,** 132–139 (2005).
20. Chawarska, K., Klin, A. & Volkmar, F. R. *Autism Spectrum Disorders in Infants and Toddlers: Diagnosis, Assessment and Treatment* (Guilford Press, 2008).
21. Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T. & Klin, A. Autism and pervasive developmental disorders. *J. Child Psychol. Psychiatry* **45,** 135–170 (2004).
22. Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* **59,** 809–816 (2002).
23. Jones, W., Carr, K. & Klin, A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-olds with autism spectrum disorder. *Arch. Gen. Psychiatry* **65,** 946–954 (2008).
24. Schultz, R. T. Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *Int. J. Dev. Neurosci.* **23,** 125–141 (2005).
25. Dalton, K. M. *et al.* Gaze fixation and the neural circuitry of face processing in autism. *Nature Neurosci.* **8,** 519–526 (2005).
26. Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L. & Stone, W. L. Visual recognition of biological motion is impaired in children with autism. *Psychol. Sci.* **14,** 151–157 (2003).
27. Pavlova, M. & Sokolov, A. Orientation specificity in biological motion perception. *Percept. Psychophys.* **62,** 889–899 (2000).
28. Reid, V. M., Hoehl, S. & Striano, T. The perception of biological motion by infants: an event-related potential study. *Neurosci. Lett.* **395,** 211–214 (2006).
29. Driver, J. & Spence, C. Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10,** R731–R735 (2000).
30. Chen, P. Y. & Popovich, P. M. *Correlation: Parametric and Nonparametric Measures* 20–23 (Sage Publications, 2002).
31. Von Collani, E. & Dräger, K. *Binomial Distribution Handbook for Scientists and Engineers* 182–187 (Birkhäuser, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the National Institutes of Mental Health (U54-MH66494), by Autism Speaks (piloting of methods), and by the Simons Foundation (current methodological developments). We wish to thank the families of the children included in this study for their time and participation. We would also like to thank F. Shic for conceptual discussions and thoughts on methods. We would like to thank L. Dickholtz, A. Kord, L. Else and M. Campisi for their assistance in creating the biological motion stimuli. We also wish to thank K. Carr, A. Blank, A. Bhatt, S. Shultz and K. Knoch for their help in this research project, and our colleagues K. Chawarska, R. Paul and F. Volkmar for conceptual discussions and for their contributions to the clinical characterization of the samples.

Author Contributions A.K. and W.J. developed the initial idea and design of the study, interpreted data, wrote the final manuscript, and take full responsibility for the integrity of data and the accuracy of data analysis. A.K. supervised participant characterization. W.J. supervised all technical aspects of experimental procedure, data acquisition and analysis. P.G. contributed to initial development of AVS methods and data analysis. D.J.L., with W.J. and G.R., developed the final AVS methods. G.R., with W.J. and A.K., helped develop new animations for the second experiment. W.J. created the figures. A.K. and W.J. performed the final revision of the manuscript for intellectual content.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.K. (ami.klin@yale.edu) or W.J. (warren.jones@yale.edu).

METHODS

Experimental procedure and setting. Throughout the procedure, toddlers were accompanied by a parent or primary caregiver. To begin the session, the child and caregiver entered the laboratory room while a children's video (for example, Baby Mozart, Elmo) played on a computer monitor. The computer monitor was mounted within a wall panel, and audio was played through a set of concealed speakers. The toddler was seated and buckled into a car seat mounted on a pneumatic lift so that the viewing height (line-of-sight) was standardized for all children. Viewers' eyes were 30 inches (76.2 cm) from the computer monitor, which subtended an approximately $23^\circ \times 30^\circ$ portion of each child's visual field. Lights in the room were dimmed so that only images shown on the computer monitor could be easily seen. The experimenter was concealed from the child's view throughout the testing session but was able to monitor the child at all times by means of an eye-tracking camera and by a second video camera that filmed a full-body image of the child.

After the child was comfortably watching a familiar children's video, the experimenter triggered the presentation of onscreen calibration targets. This was done with software that paused the playing video and presented a calibration target on the otherwise blank screen. A five-point calibration scheme was used, with a variety of small cartoon animations as well as spinning and/or flashing points of light, ranging in size from 0.5° to 1.5° of visual angle, all with accompanying sounds. The calibration routine was followed by a verification of calibration in which more animations were presented at nine on-screen locations.

Throughout the remainder of the testing session, animated targets (as used in the calibration process) were shown between experimental videos to measure drift in calibration. In this way, accuracy of the eye-tracking data was verified before beginning the experimental trials and was then repeatedly checked between video segments as the testing continued. In the case that drift exceeded 3°, data collection was stopped and the child was re-calibrated before further videos were presented.

All aspects of the experimental protocol were performed by personnel 'blinded' to the diagnostic status of child. Most aspects of the data acquisition and all aspects of coding, processing, and data summary are automated so that the separation between the diagnostic characterization protocol and the experimental protocol is assured.

Motion capture stimuli and preferential viewing. Point-light biological motion animations were shown as full-screen audiovisual stimuli on a 20-inch (50.8-cm) computer monitor (refresh rate of 60 Hz non-interlaced). Video frames were 8-bit greyscale images, 640×480 pixels in resolution. The video frame rate of presentation was 30 frames per s. The audio track was a single (mono) channel sampled at 44.1 kHz. The duration of each animation varied with the content of the action, with a mean duration of 30.5 s and a range of 26.4 to 35.5 s. A centring cue lasting 2,800 ms was played immediately before the start of presentation of each animation.

The animations were created with a process called motion capture, in which three-dimensional representations of live performances are recorded in real-time from the movements of actors. The stimuli were created with equipment and support from Animazoo and MetaMotion. Motion is recorded (in three planes of space) directly into computer files by means of an electronic suit worn by the actor. The suit has potentiometers at each joint in the body that track and record movements of the individual wearing the suit. This method enabled us to create a variety of stimuli tailored to young children, featuring routines relevant to childhood experience. As noted, there were five point-light animations portraying an adult's attempts to engage a child. They included the following social approaches: (1) getting the child's attention, (2) playing peek-a-boo ('I can't see you'), (3) playing with a teddy bear, (4) playing pat-a-cake, and (5) enacting a feeding routine.

Preferential viewing in our design was a binary choice, upright versus inverted. To determine viewing preferences that were significantly different from chance, we modelled total viewing time as a binomial distribution. The average viewing time per participant, in the number of frames of video fixated by the toddlers was 5,827 total, 1,165 per animation type. Modelling the binary outcome for this number of frames indicates that results between 47% and 53% should be considered random viewing[31].

Data acquisition and analysis. As noted, visual scanning was measured with eye-tracking equipment (ISCAN, Inc.). The equipment uses a dark pupil/corneal reflection technique with data collected at the rate of 60 Hz (double the frequency of stimuli presentation and of sufficient resolution to identify on- and offset of saccades at a threshold rotational velocity of 30° per s[32]). The eye-tracking camera was mounted remotely, concealed from the child's view behind an infrared filter in a wall panel.

Analysis of eye movements and coding of preferential fixation data were performed with software written in MATLAB. The first phase of analysis was an automated identification of blinks, saccades, and off-screen fixations. Saccades were identified by a velocity threshold. Blinks were identified by eyelid closure (via the rate of change of pupil size and by change in vertical centre-of-pupil data). Off-screen fixations, when a child looked away from the presentation screen, were identified by pupil minus corneal reflection vectors mapping to locations beyond the screen bounds. In the second phase of analysis, eye movements identified as fixations were coded relative to the upright and inverted animations (Fig. 1b–g and Supplementary Movies 1–3).

From within the 304.7 s of total viewing data (9,142 video frames), non-fixation data were not significantly different between the three groups (ANOVA): for all non-fixation data (saccades + blinks + off-screen), ASD = 35.8% (s.d. = 16.4), typically developing = 35.2% (16.1), developmentally delayed = 37.8% (13.2), $F_{2,73} = 0.15$, $P = 0.860$; or separately for saccades, ASD = 15.2% (7.7), typically developing = 13.1% (4.2), developmentally delayed = 15.4% (7.0), $F_{2,73} = 1.3$, $P = 0.277$; for blinks, ASD = 7.4 (7.8), typically developing = 3.9 (5.4), developmentally delayed = 4.7 (5.2), $F_{2,73} = 2.2$, $P = 0.113$; or for off-screen fixations, ASD = 13.3% (12.4), typically developing = 18.3% (13.7), developmentally delayed = 17.8% (12.6), $F_{2,73} = 1.05$, $P = 0.355$.

Quantification of AVS. To quantify AVS, we tracked the locations of the point-lights in our stimuli and compared the change in their motion with the change in the animation's audio soundtrack. Related methods have been described previously[33,34].

We measured the spatial trajectories (x-y location over time) of all point-lights throughout each biological motion animation: 16 point-lights each for the upright and inverted sides of the animation, across five animations (counterbalanced presentations necessarily yielded identical location data, just reversed for left or right presentation). We stored the locations of the point-lights at each frame in each of the animations as a matrix of size $N \times 2 \times 16$, in which the rows (N) signified frames, the columns (2) signified (x, y) coordinate location data, and the Z dimension (16) signified each individual point-light on one side of the animation screen. On the basis of the manner in which the stimuli were created, the location data of the inverted point-light objects were identical to the location data of the upright point-lights except that they were inverted in space and reversed in time.

From each point-light's trajectory, we calculated its velocity over time, and then its change in velocity, $|\Delta v|$. We smoothed the change in velocity data with a moving-average window-size of three samples. This signal, for each of the 32 point-lights in a given animation (16 upright, 16 inverted) provided our measure of change in motion.

To measure the change in audio over time, we measured the audio amplitude of the soundtrack (its short-term amplitude envelope) and then calculated its rate of change, $|\Delta A|$. The short-term amplitude envelope (SAE) of the audio track was computed as the root mean squared (r.m.s.) of a 100-ms square wave moving average of the original audio signal[35]. To normalize this signal for global variance in intensity, we computed two filtered versions of the SAE: one filtered with a moving-average square window of seven samples (local window), and a second with a square window of 35 samples (global window). We then divided the signal filtered at the local window by the signal filtered at the global window. This step is included to normalize for global variance in intensity of a signal while preserving local signal change[36].

Having calculated both change in audio and change in motion for each animation and for all point-lights, we then computed our measure of AVS. By multiplying the change in motion data (each point light, 16 upright, 16 inverted per animation) by the change in audio data (one signal per animation), we generated an audiovisual coincidence matrix for each animation: this gives an AVS value for each point-light at each point in time. High values indicate increased change in motion occurring synchronously with increased change in audio amplitude. Conversely, low values indicate either that one signal was low while the other was high (so the two were not changing synchronously with one another), or that both signals were low (so that with no movement and no change in audio, there was little observable AVS).

To map the AVS values (computed for each point-light) back into the visual space of each presented animation, we overlaid our computed AVS data onto the locations of the point-lights in the original animations (for example, see Supplementary Movie 3). In the movie, colour data are scaled to the maximum AVS value. For each frame of each animation, this generates a matrix of AVS depicting the amount of AVS at each pixel at each frame.

To quantify AVS over the entire duration of an animation, we summed all frames of AVS data, yielding a cumulative map of AVS for each movie. We smoothed these cumulative maps with an averaging filter of size [10 10]. Filter sizes of 6, 8, 10, 12, 16 and 20 all gave similar results. A plot of cumulative AVS is shown in Fig. 3h. AVS level is in arbitrary units (change in motion multiplied by change in audio, summed over all frames), and the maximum value of AVS depends on the number of frames in a given animation (that is, an animation with 1,000 frames is likely to have a larger cumulative signal than one with only 800 frames; to normalize for comparison across animations, we converted to percentages as described later).

To compare AVS on the upright side versus the inverted side, we found the maximum cumulative AVS on each side (for example, 600 on the upright, 400 on the inverted), and converted these values to percentage of total: 600/(600 + 400) = 60%, 400/(600 + 400) = 40%. We then computed a difference ratio of upright to inverted AVS (as plotted in Fig. 4) as the upright percentage minus the inverted percentage: 60% − 40% = 20% (0.2 on the plot in Fig. 4). This generated a normalized score comparable across animations that could be used as a predictor of viewers' looking patterns. We could then test whether or not preferential viewing in our stimuli was related to the level of AVS, as the relative levels of synchrony on the upright versus inverted side provide predictions of expected viewing behaviour.

32. Leigh, R. J. & Zee, D. S. *The Neurology of Eye Movements* 3rd edn 94 (Oxford Univ. Press, 1999).
33. Hershey, J. & Movellan, J. R. in *Advances in Neural Information Processing Systems 12* (eds Solla, S. A., Leen, T. K. & Muller, K. R.) 813–819 (MIT Press, 2000).
34. Bredin, H. & Chollet, G. Measuring audio and visual speech synchrony: methods and applications. *IET International Conference on Visual Information Engineering* 255–260 (2006).
35. Rabiner, L. R. & Schafer, R. W. *Digital Processing of Speech Signals* (Prentice-Hall, 1978).
36. Tchorz, J. & Kollmeier, B. Estimation of the signal-to-noise ratio with amplitude modulation spectrograms. *Speech Commun.* **38,** 1–17 (2002).

Vol 459|14 May 2009|doi:10.1038/nature07935

LETTERS

Single Lgr5 stem cells build crypt–villus structures *in vitro* without a mesenchymal niche

Toshiro Sato[1], Robert G. Vries[1], Hugo J. Snippert[1], Marc van de Wetering[1], Nick Barker[1], Daniel E. Stange[1], Johan H. van Es[1], Arie Abo[2], Pekka Kujala[3], Peter J. Peters[3] & Hans Clevers[1]

The intestinal epithelium is the most rapidly self-renewing tissue in adult mammals. We have recently demonstrated the presence of about six cycling Lgr5$^+$ stem cells at the bottoms of small-intestinal crypts[1]. Here we describe the establishment of long-term culture conditions under which single crypts undergo multiple crypt fission events, while simultanously generating villus-like epithelial domains in which all differentiated cell types are present. Single sorted Lgr5$^+$ stem cells can also initiate these crypt–villus organoids. Tracing experiments indicate that the Lgr5$^+$ stem-cell hierarchy is maintained in organoids. We conclude that intestinal crypt–villus units are self-organizing structures, which can be built from a single stem cell in the absence of a non-epithelial cellular niche.

The self-renewing epithelium of the small intestine is ordered into crypts and villi[2]. Cells are newly generated in the crypts and are lost by apoptosis at the tips of the villi, with a turnover time of 5 days in the mouse. Self-renewing stem cells have long been known to reside near the crypt bottom and to produce the rapidly proliferating transit amplifying (TA) cells. The estimated number of stem cells is between four and six per crypt. Enterocytes, goblet cells and enteroendocrine cells develop from TA cells and continue their migration in coherent bands along the crypt–villus axis. The fourth major differentiated cell type, the Paneth cell, resides at the crypt bottom. We have recently identified a gene, *Lgr5*, that is specifically expressed in cycling crypt base columnar (CBC) cells that are interspersed between the Paneth cells[1]. Using a mouse in which a green fluorescent protein (GFP)/tamoxifen-inducible *Cre* recombinase cassette was integrated into the *Lgr5* locus, we showed by lineage tracing that the Lgr5$^+$ cells constitute multipotent stem cells that generate all cell types of the epithelium[1], even when assessed 14 months after induction of Cre[3].

Although a variety of culture systems have been described[4–7], no long-term culture system has been established that maintains basic crypt–villus physiology[2]. We attempted to design such a culture system by combining previously defined insights in the growth requirements of intestinal epithelium. First, Wnt signalling is a pivotal requirement for crypt proliferation[8–10] and the Wnt agonist R-spondin 1 induces marked crypt hyperplasia *in vivo*[11]. Second, signalling by epidermal growth factor (EGF) is associated with intestinal proliferation[12]. Third, transgenic expression of Noggin induces an expansion of crypt numbers[13]. Fourth, isolated intestinal cells undergo anoikis outside the normal tissue context[14]. Because laminin (α1 and α2) is enriched at the crypt base[15], we explored the use of laminin-rich Matrigel to support intestinal epithelial growth. Matrigel-based cultures have been used successfully for the growth of mammary epithelium[16].

Mouse crypt preparations were suspended in Matrigel. Crypt growth required EGF and R-spondin 1 (Supplementary Fig. 1a). Passaging revealed a requirement for Noggin (Supplementary Fig. 1b). The cultured crypts behaved in a stereotypical manner (Fig. 1a; Supplementary Movie 1). The upper opening rapidly became sealed, and the lumen filled with apoptotic cells. The crypt region underwent continuous budding events, reminiscent of crypt fission[17]. Paneth cells were always present at the bud site. Most crypts could be cultured (Fig. 1b). Further expansion created organoids, comprising more than 40 crypt domains surrounding a central lumen lined by a villus-like epithelium ('villus domain') (Fig. 1c–e). Staining with E-cadherin revealed a single cell layer (Supplementary Fig. 2). At weekly intervals, organoids were mechanically dissociated and replated at one-fifth of the pre-plating density. Organoids were cultured for more than 8 months without losing the characteristics described below. Expression analysis by microarray revealed that organoids remained highly similar to freshly isolated small-intestinal crypts, for instance when compared with fresh colon crypts (Supplementary Fig. 3). Moreover, no significant induction of stress-related genes was observed (Supplementary Table 1).

Culture of *Lgr5–EGFP–ires–CreERT2* crypts revealed Lgr5–GFP$^+$ stem cells intermingled with Paneth cells at the crypt base. Wnt activation, as demonstrated by nuclear β-catenin (Supplementary Figs 4a and 9) and expression of the Wnt target genes *Lgr5* (Fig. 1d) and *EphB2* (ref. 18) (Supplementary Fig. 4b), was confined to the crypts. Apoptotic cells were shed into the central lumen, a process reminiscent of the shedding of apopotic cells at villus tips *in vivo* (Supplementary Fig. 4c). Metaphase spreads of organoids more than 3 months old consistently revealed 40 chromosomes in each cell ($n = 20$) (Supplementary Fig. 4d). We found no evidence for the presence of myofibroblasts or other non-epithelial cells (Supplementary Fig. 5).

We cultured crypts from *Lgr5–EGFP–ires–CreERT2* mice crossed with the *Cre*-activatable *Rosa26–LacZ* reporter to allow lineage tracing. Directly after induction with low-dose tamoxifen, we noted single labelled cells (Supplementary Fig. 4e, g). More than 90% of these generated entirely blue crypts (Supplementary Fig. 4e–g), implying that the Lgr5–GFP$^+$ cells did indeed retain stem cell properties. Crypts from the *Cre*-activatable *Rosa26–YFP* reporter[19,20] mouse allowed lineage tracing by confocal analysis. Directly after treatment with tamoxifen, we noted single labelled cells that induced lineage tracing over the following days, both in freshly isolated crypts (Supplementary Fig. 6a–c) and in established organoids (Supplementary Fig. 6d). Supplementary Movie 2 represents four days of lineage tracing, revealing green Lgr5$^+$ cells and YFP$^+$ offspring (pseudocolour red) against the backdrop of a growing organoid.

Recently, mammary gland epithelial structures were established from single stem cells *in vitro*[21]. When single Lgr5–GFP^{hi} cells were sorted, these died immediately. The Rho kinase inhibitor Y-27632, which inhibits anoikis of embryonic stem cells[22], significantly

[1]Hubrecht Institute and University Medical Center Utrecht, Uppsalalaan 8, 3584CT Utrecht, The Netherlands. [2]Nuvelo, Inc., 201 Industrial Road, Suite 310, San Carlos, California 94070-6211, USA. [3]The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

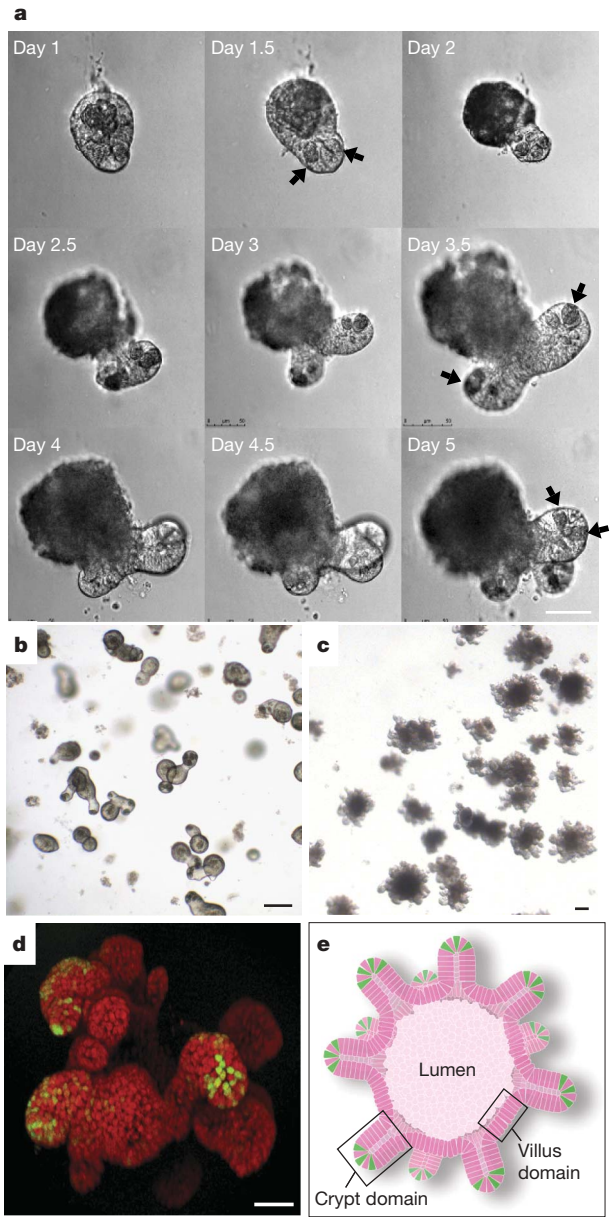


Figure 1 | Establishment of intestinal crypt culture system. **a**, Time course of an isolated single crypt growth. Differential interference contrast image reveals granule-containing Paneth cells at crypt bottoms (arrows). **b**, **c**, Single isolated crypts efficiently form large crypt organoids within 14 days; **b**, on day 5; **c**, on day 14. **d**, Three-dimensional reconstructed confocal image after 3 weeks in culture. Lgr5–GFP^{+} stem cells (green) are localized at the tip of crypt-like domains. Counterstain, ToPro-3 (red). **e**, Schematic representation of a crypt organoid, consisting of a central lumen lined by villus-like epithelium and several surrounding crypt-like domains. Scale bar, 50 μm.

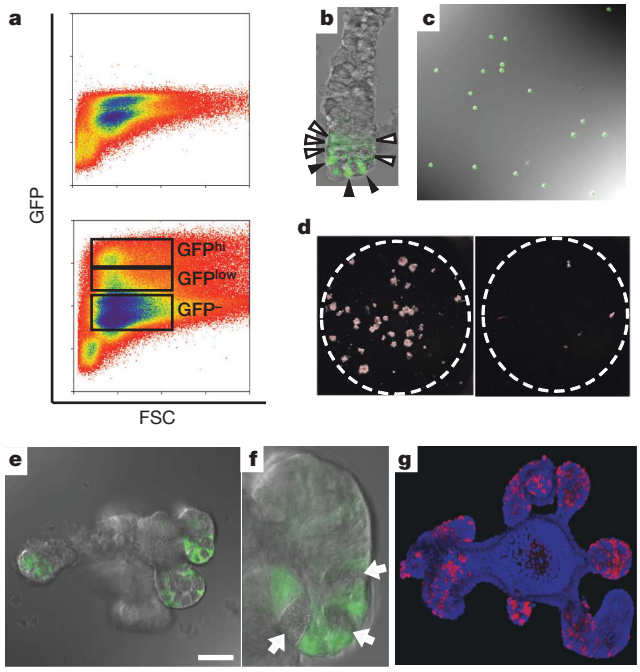


Figure 2 | Single Lgr5^{+} cells generate crypt–villus structures. **a**, Lgr5–GFP^{+} cells from an *Lgr5–EGFP–ires–CreERT2* intestine (bottom); wild-type cells (top). Two positive populations, GFP^{hi} and GFP^{low}, are discriminated. FSC, forward scatter. **b**, Confocal analysis of a freshly isolated crypt. Black arrowheads, GFP^{hi}; white arrowheads, GFP^{low}. **c**, Sorted GFP^{hi} cells. **d**, 1,000 sorted GFP^{hi} cells (left) and GFP^{low} cells (right) after 14 days in culture. **e**, **f**, Fourteen days after sorting, single GFP^{hi} cells form crypt organoids, with Lgr5–GFP^{+} cells and Paneth cells (white arrows) located at crypt bottoms. Scale bar, 50 μm. **f**, Higher magnification of **e**. **g**, Organoids cultured with the thymidine analogue EdU (red) for 1 h. Note that only crypt domains incorporate EdU. Counterstain, 4,6-diamidino-2-phenylindole (DAPI; blue).

decreased this cell death. Because cell-to-cell Notch signalling is essential to maintain proliferative crypts[23], we also provided a Notch-agonistic peptide[24]. Under these conditions, significant numbers of Lgr5–GFP^{hi} cells survived and formed large crypt organoids. Organoids formed rarely when GFP^{low} daughter cells were seeded (Fig. 2d). Multiple Lgr5–GFP^{hi} cells were intermingled with Paneth cells at crypt bottoms (Fig. 2e, f). Incorporation of 5-ethynyl-2′-deoxyuridine (EdU, a thymidine analogue) revealed S-phase cells in the crypts (Fig. 2g).

We sorted cells at one cell per well, visually verified the presence of single cells and followed the resulting growth. In each of four individual experiments, we identified and followed 100 single cells. On average, about 6% of the Lgr5–GFP^{hi} cells grew out into organoids, whereas the remaining cells typically died within the first 12 h, presumably as a result of physical and/or biological stress inherent in the isolation procedure. GFP^{low} cells rarely grew out (Fig. 3a). Figure 3b and Supplementary Fig. 7 illustrate the growth of an organoid from a single Lgr5–GFP^{hi} cell. By four days of culture, the structures consisted of about 100 cells, which is consistent with the 12-h cell cycle of proliferative crypt cells[25] (Fig. 3c). After 2 weeks, the organoids were dissociated into single cells and replated to form new organoids (Fig. 3d). This procedure could be repeated at least four times on a two-weekly basis, without apparent loss of replating efficiency.

The organoids derived from single stem cells were indistinguishable in appearance from those derived from whole crypts. Paneth cells and stem cells were located at crypt bottoms (Figs 2e, f and 4c, g). Fully polarized enterocytes, as demonstrated by villin^{+} mature brush borders and apical alkaline phosphatase, lined the central lumen (Fig. 4a, e, i). Goblet cells (Muc2^{+}, Fig. 4b; periodic acid–Schiff (PAS)$^{+}$, Fig. 4f) and enteroendocrine cells (chromogranin A^{+}, Fig. 4d; synaptophysin^{+}, Fig. 4h) were scattered throughout the organoid structure. Four types of mature cell were recognized by electron microscopy (Fig. 4i–l). Non-epithelial (stromal/mesenchymal) cells were absent, an observation confirmed by electron-microscopic imaging (Fig. 4i–p and Supplementary Fig. 8c–g). Both the crypts (Fig. 4m–o) and the central luminal epithelium (Fig. 4p) consisted of a single layer of polarized epithelial cells resting directly on the Matrigel support. High-resolution images of these electron-microscopic pictures are given in Supplementary Fig. 9. We frequently noted small intercellular vacuoles, possibly an indicator of culture-induced or fixation-induced stress (Fig. 4i–p and Supplementary Fig. 8).

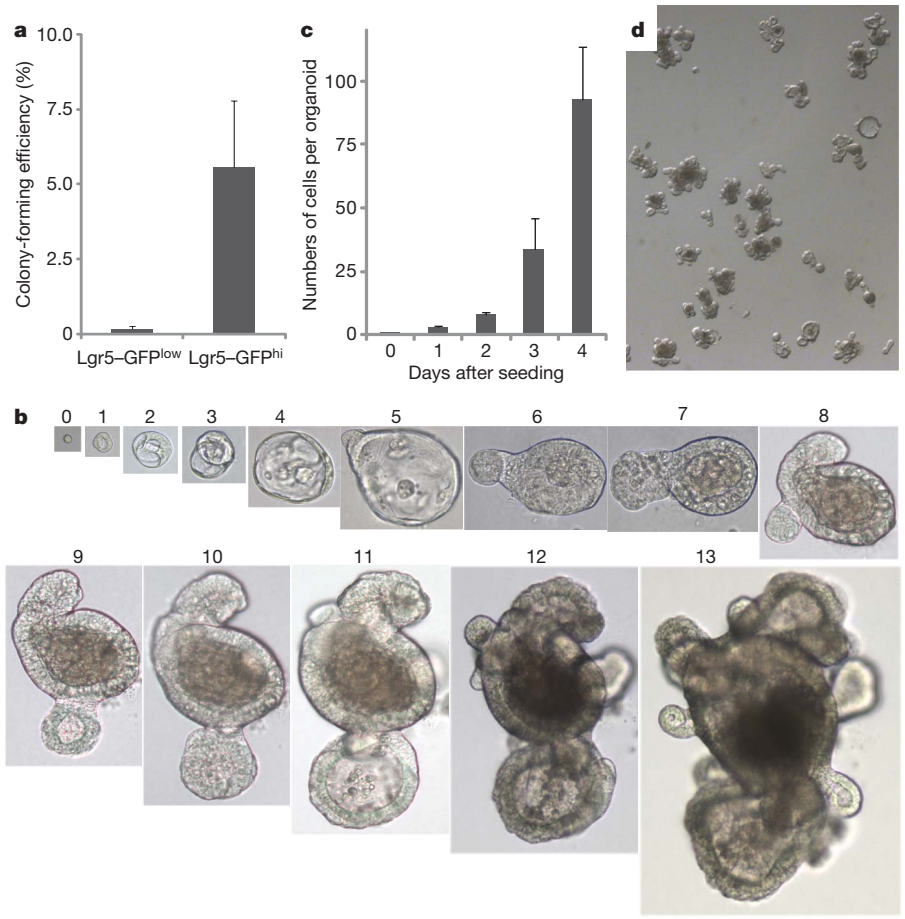
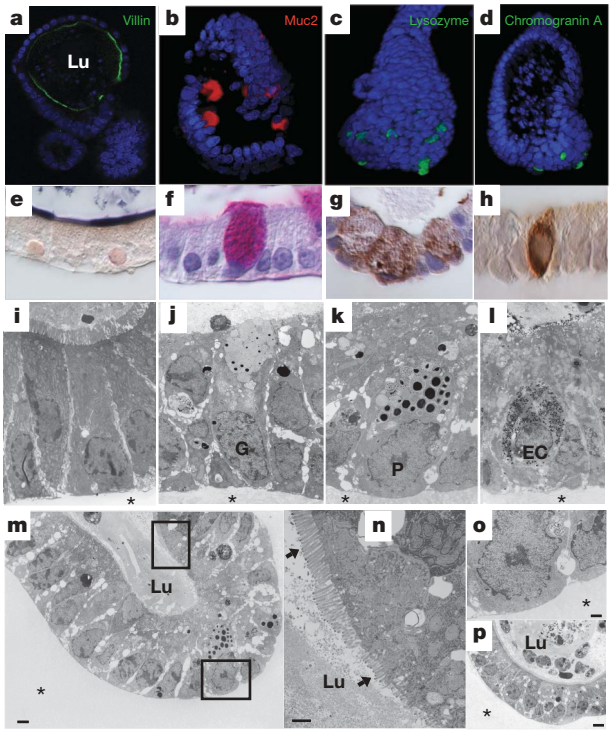


Figure 3 | Colony-forming efficiency of single cells sorted in individual wells. **a**, Colony-forming efficiency was calculated from 100 single sorted GFP^{hi} cells. **b**, An example of a successfully growing single GFP^{hi} cell. Numbers above the images are the days of growth. **c**, Numbers of cells per single organoid averaged for five growing organoids. **d**, A single-cell suspension derived from a single-cell-derived-organoid was replated and grown for 2 weeks. Error bars in **c** and **d** indicate s.e.m. Original magnifications in **b**: days 0–4, ×40; days 5–7, ×20; days 8–11, ×10; days 12 and 13, ×4.



It is well known that epithelial crypts are in intimate contact with subepithelial myofibroblasts[26–28], and it is generally believed that the latter cells create a specialized cellular niche at crypt bottoms[27,29,30]. Such a niche would create a unique environment to anchor and support the intestinal stem cells. We now show that a self-renewing epithelium can be established by a limited set of growth signals that are uniformly presented. Despite this, the isolated stem cells autonomously generate asymmetry in a highly stereotypical fashion. This rapidly leads to the formation of crypt-like structures with *de novo* generated stem cells and Paneth cells located at their bottoms and filled with TA cells. These crypt-like structures feed into villus-like luminal domains consisting of postmitotic enterocytes, in which apoptotic cells pinch off into the lumen in a manner reminiscent of cell loss at villus tips. The paradoxical observation that single cells exposed to a uniform growth-promoting environment can generate asymmetric structures is particularly evident

Figure 4 | Composition of single stem cell-derived organoids. **a**–**d**, Confocal image for villin (**a**, green, enterocytes), Muc2 (**b**, red, goblet cells), lysozyme (**c**, green, Paneth cells) and chromogranin A (**d**, green, enteroendocrine cells). Counterstain, DAPI (blue). **e**–**h**, Paraffin sections stained for alkaline phosphatase (**e**, green, enterocytes), periodic acid-Schiff (**f**, red, goblet cells), lysozyme (**g**, brown, Paneth cells) and synaptophysin (**h**, brown, enteroendocrine cells). **i**–**p**, Electron microscopy demonstrates enterocytes (**i**), goblet cells (**j**), Paneth cells (**k**) and enteroendocrine cells (**l**). **m**–**o**, Low-power crypt images. **n**, **o**, Higher magnifications of **m**. **n**, Maturation of brush border (black arrows). **p**, Low-power villus domain image. Lu, lumen with apoptotic bodies, lined by polarized enterocytes. G, goblet cells; EC, enteroendocrine cells; P, Paneth cells; *, Matrigel. Scale bars, 5 μm (m, p) and 1 μm (n, o).

on scrutiny of the Wnt pathway. Although all cells are exposed to R-spondin 1, only cells in crypts display hallmarks of active Wnt signalling; that is, nuclear β-catenin and the expression of Wnt target genes. Apparently, differential responsiveness to Wnt signalling rather than differential exposure to extracellular Wnt signals lies at the heart of the formation of a crypt–villus axis.

We conclude that a single $Lgr5^+$ intestinal stem cell can operate independently of positional cues from its environment and that it can generate a continuously expanding, self-organizing epithelial structure reminiscent of normal gut. The culture system described will simplify the study of stem-cell-driven crypt–villus biology. Moreover, it may open up new avenues for regenerative medicine and gene therapy.

METHODS SUMMARY

Mice. Outbred mice 6–12 weeks old were used. Generation and genotyping of the *Lgr5–EGFP–Ires–CreERT2* allele[1] has been described previously[3]. Rosa26–lacZ or YFP–Cre reporter mice were obtained from Jackson Labs.

Crypt isolation, cell dissociation and cell culture. Crypts were released from murine small intestine by incubation for 30 min at 4 °C in PBS containing 2 mM EDTA (Supplementary Methods). Isolated crypts were counted and pelleted. A total of 500 crypts were mixed with 50 μl of Matrigel (BD Bioscience) and plated in 24-well plates. After polymerization of Matrigel, 500 μl of crypt culture medium (Advanced DMEM/F12 (Invitrogen)) containing growth factors (10–50 ng ml^{-1} EGF (Peprotech), 500 ng ml^{-1} R-spondin 1 (ref. 11) and 100 ng ml^{-1} Noggin (Peprotech)) was added. For sorting experiments, isolated crypts were incubated in culture medium for 45 min at 37 °C, followed by trituration with a glass pipette. Dissociated cells were passed through cell strainer with a pore size of 20 μm. GFP^{hi}, GFP^{low} and GFP^- cells were sorted by flow cytometry (MoFlo; Dako). Single viable epithelial cells were gated by forward scatter, side scatter and pulse-width parameter, and by negative staining for propidium iodide. Sorted cells were collected in crypt culture medium and embedded in Matrigel containing Jagged-1 peptide (1 μM; AnaSpec) at 1 cell per well (in 96-well plates, 5 μl Matrigel). Crypt culture medium (250 μl for 48-well plates, 100 μl for 96-well plates) containing Y-27632 (10 μM) was overlaid. Growth factors were added every other day and the entire medium was changed every 4 days. For passage, organoids were removed from Matrigel and mechanically dissociated into single-crypt domains, and then transferred to fresh Matrigel. Passage was performed every 1–2 weeks with a 1:5 split ratio.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July 2008; accepted 24 February 2009.
Published online 29 March 2009.

1. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
2. Bjerknes, M. & Cheng, H. Intestinal epithelial stem cells and progenitors. *Methods Enzymol.* **419,** 337–383 (2006).
3. Barker, N., van de Wetering, M. & Clevers, H. The intestinal stem cell. *Genes Dev.* **22,** 1856–1864 (2008).
4. Evans, G. S., Flint, N., Somers, A. S., Eyden, B. & Potten, C. S. The development of a method for the preparation of rat intestinal epithelial cell primary cultures. *J. Cell Sci.* **101,** 219–231 (1992).
5. Whitehead, R. H., Demmler, K., Rockman, S. P. & Watson, N. K. Clonogenic growth of epithelial cells from normal colonic mucosa from both mice and humans. *Gastroenterology* **117,** 858–865 (1999).
6. Fukamachi, H. Proliferation and differentiation of fetal rat intestinal epithelial cells in primary serum-free culture. *J. Cell Sci.* **103,** 511–519 (1992).
7. Perreault, N. & Jean-Francois, B. Use of the dissociating enzyme thermolysin to generate viable human normal intestinal epithelial cell cultures. *Exp. Cell Res.* **224,** 354–364 (1996).
8. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19,** 379–383 (1998).
9. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17,** 1709–1713 (2003).
10. Kuhnert, F. *et al.* Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl Acad. Sci. USA* **101,** 266–271 (2004).
11. Kim, K. A. *et al.* Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* **309,** 1256–1259 (2005).
12. Dignass, A. U. & Sturm, A. Peptide growth factors in the intestine. *Eur. J. Gastroenterol. Hepatol.* **13,** 763–770 (2001).
13. Haramis, A. P. *et al. De novo* crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* **303,** 1684–1686 (2004).
14. Hofmann, C. *et al.* Cell–cell contacts prevent anoikis in primary human colonic epithelial cells. *Gastroenterology* **132,** 587–600 (2007).
15. Sasaki, T., Giltay, R., Talts, U., Timpl, R. & Talts, J. F. Expression and distribution of laminin α1 and α2 chains in embryonic and adult mouse tissues: an immunochemical approach. *Exp. Cell Res.* **275,** 185–199 (2002).
16. Stingl, J., Eaves, C. J., Zandieh, I. & Emerman, J. T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res. Treat.* **67,** 93–109 (2001).
17. St Clair, W. H. & Osborne, J. W. Crypt fission and crypt number in the small and large bowel of postnatal rats. *Cell Tissue Kinet.* **18,** 255–262 (1985).
18. Batlle, E. *et al.* β-Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111,** 251–263 (2002).
19. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1,** 4 (2001).
20. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21,** 70–71 (1999).
21. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439,** 993–997 (2006).
22. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnol.* **25,** 681–686 (2007).
23. van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435,** 959–963 (2005).
24. Li, L. *et al.* The human homolog of rat Jagged1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1. *Immunity* **8,** 43–55 (1998).
25. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. *Am. J. Anat.* **141,** 461–479 (1974).
26. Powell, D. W. *et al.* Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *Am. J. Physiol.* **277,** C183–C201 (1999).
27. Yen, T. H. & Wright, N. A. The gastrointestinal tract stem cell niche. *Stem Cell Rev.* **2,** 203–212 (2006).
28. Kedinger, M. *et al.* Intestinal epithelial–mesenchymal cell interactions. *Ann. NY Acad. Sci.* **859,** 1–17 (1998).
29. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414,** 98–104 (2001).
30. Li, L. & Xie, T. Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* **21,** 605–631 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. van den Born, J. Korving, H. Begthel and S. van den Brink for technical assistance, and N. Ong and M. van den Bergh Weerman for technical assistance with electron microscopy.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C. (h.clevers@niob.knaw.nl).

METHODS

Reagents. Murine recombinant EGF and Noggin were purchased from Peprotech. Human recombinant R-spondin 1 (ref. 11), Y-27632 (Sigma), 4-hydroxytamoxifen (Sigma) and EdU (Invitrogen) were used for culture experiments. The following antibodies were used for immunostaining: anti-lysozyme (Dako), anti-Synaptophysin (Dako), anti-bromodeoxyuridine (Roche), anti-β-catenin (BD Bioscience), anti-E-cadherin (BD Bioscience), anti-smooth muscle actin (Sigma), anti-EphB2 and anti-EphB3 (R&D), anti-villin, anti-Muc2 and anti-chromogranin A (Santa Cruz) and anti-caspase-3 (Cell Signaling).

Crypt isolation. Isolated small intestines were opened longitudinally, and washed with cold PBS. The tissue was chopped into around 5 mm pieces, and further washed with cold PBS. The tissue fragments were incubated in 2 mM EDTA with PBS for 30 min on ice. After removal of EDTA medium, the tissue fragments were vigorously suspended by using a 10-ml pipette with cold PBS. The supernatant was the villous fraction and was discarded; the sediment was resuspended with PBS. After further vigorous suspension and centrifugation, the supernatant was enriched for crypts. This fraction was passed through a 70-μm cell strainer (BD Bioscience) to remove residual villous material. Isolated crypts were centrifuged at 150–200*g*. for 3 min to separate crypts from single cells. The final fraction consisted of essentially pure crypts and was used for culture or single cell dissociation.

Tamoxifen induction and staining with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). To activate CreERT2, crypts were incubated with a low dose of 4-hydroxytamoxifen (100 nM) for 12 h and cultured in crypt culture medium. X-Gal staining was performed as described previously[3]. No staining was seen without 4-hydroxytamoxifen treatment.

Electron microscopic analysis. As described previously[3], Matrigel including crypt organoids was fixed in Karnovsky's fixative (2% paraformaldehyde, 2.5% glutaraldehyde, 0.1 M sodiumm cacodylate, 2.5 mM $CaCl_2$, 5 mM $MgCl_2$, pH 7.4) for 5 h at room temperature (18–22 °C). The samples were embedded in Epon resin and were examined with a Phillips CM10 microscope.

Microarray analysis: gene expression analysis of colonic crypts, small-intestinal crypts and organoids. Freshly isolated small-intestinal crypts from two mice were divided into two parts. RNA was directly isolated from one part (RNeasy Mini Kit; Qiagen); the other part was cultured for 1 week, followed by RNA isolation. We prepared labelled antisense RNA in accordance with the manufacturer's instructions (Agilent Technologies). Differentially labelled cRNA from small-intestinal crypts and organoids were hybridized separately for the two mice on a 4 × 44k Agilent Whole Mouse Genome Dual Colour Microarray (G4122F) in two dye-swap experiments, resulting in four individual arrays. Additionally, isolated colonic crypts were hybridized against differentially labelled small-intestinal crypts in two dye-swap experiments, resulting in four individual arrays. Microarray signal and background information were retrieved with Feature Extraction (v. 9.5.3; Agilent Technologies). All data analyses were performed with ArrayAssist (5.5.1; Stratagene, Inc.) and Microsoft Excel (Microsoft Corporation). Raw signal intensities were corrected by subtracting local background. Negative values were changed into a positive value close to zero (standard deviation of the local background) to permit the calculation of ratios between intensities for features present in only one channel. Normalization was performed by applying a locally weighted linear regression (LOWESS) algorithm, and individual features were filtered if both intensities were changed or less than double the background signal. Furthermore, non-uniform features were filtered. Data are available at GEO (Gene Expression Omnibus, accession number GSE14594). Unsupervised hierarchical clustering was performed on normalized intensities (processed signal in feature extraction) of small-intestinal or colonic crypts and organoids using Cluster 3 (distance, city block; correlation, average linkage) and visualized with TreeView. Genes were considered significantly changed if they were consistently in all arrays more than threefold enriched in organoids or crypts.

Image analysis. The images of crypt organoids were taken by either confocal microscopy with a Leica SP5, an inverted microscope (Nikon DM-IL) or a stereomicroscope (Leica, MZ16-FA). For immunohistochemistry, samples were fixed with 4% paraformaldehyde (PFA) for 1 h at room temperature, and paraffin sections were processed with standard techniques[3]. Immunohistochemistry was performed as described previously[3]. For whole-mount immunostaining, crypt organoids were isolated from Matrigel using Dispase (Invitrogen), and fixed with 4% PFA, followed by permeabilization with 0.1% Triton X-100. EdU staining followed the manufacturer's protocol (Click-IT; Invitrogen). DNA was stained with DAPI or ToPro-3 (Molecular Probes). Three-dimensional images were acquired with confocal microscopy and reconstructed with Volocity Software (Improvision).

Vol 459|14 May 2009|doi:10.1038/nature08055

LETTERS

Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column

Yanmei Shi[1], Gene W. Tyson[1] & Edward F. DeLong[1,2]

Microbial gene expression in the environment has recently been assessed via pyrosequencing of total RNA extracted directly from natural microbial assemblages. Several such 'metatranscriptomic' studies[1,2] have reported that many complementary DNA sequences shared no significant homology with known peptide sequences, and so might represent transcripts from uncharacterized proteins. Here we report that a large fraction of cDNA sequences detected in microbial metatranscriptomic data sets are comprised of well-known small RNAs (sRNAs)[3], as well as new groups of previously unrecognized putative sRNAs (psRNAs). These psRNAs mapped specifically to intergenic regions of microbial genomes recovered from similar habitats, displayed characteristic conserved secondary structures and were frequently flanked by genes that indicated potential regulatory functions. Depth-dependent variation of psRNAs generally reflected known depth distributions of broad taxonomic groups[4], but fine-scale differences in the psRNAs within closely related populations indicated potential roles in niche adaptation. Genome-specific mapping of a subset of psRNAs derived from predominant planktonic species such as *Pelagibacter* revealed recently discovered as well as potentially new regulatory elements. Our analyses show that metatranscriptomic data sets can reveal new information about the diversity, taxonomic distribution and abundance of sRNAs in naturally occurring microbial communities, and indicate their involvement in environmentally relevant processes including carbon metabolism and nutrient acquisition.

Microbial sRNAs are untranslated short transcripts that generally reside within intergenic regions (IGRs) on microbial genomes, typically ranging from 50 to 500 nucleotides in length[3]. Most microbial sRNAs function as regulators, and many are known to regulate environmentally significant processes including amino acid and vitamin biosynthesis[5], quorum sensing[6] and photosynthesis[7]. Because the identification and characterization of microbial regulatory sRNAs has relied primarily on a few model microorganisms[8–10], relatively little is known about the broader diversity and ecological relevance of sRNAs in natural microbial communities.

During a microbial gene expression study comparing four metatranscriptomic data sets from a microbial community depth profile (25 m, 75 m, 125 m and 500 m at Hawaii Ocean Time-series (HOT) Station ALOHA[11]), we discovered that a large fraction of cDNA sequences could not be assigned to protein-coding genes or ribosomal RNAs (Fig. 1). However, >28% of these unassigned cDNA reads from each data set mapped with high nucleotide identity (≥85%) to IGRs on the genomes of marine planktonic microorganisms (Supplementary Fig. 1), indicating that they may be sRNAs. Consistent with the genomic location of known sRNAs[12], many of these reads mapped on IGRs distant from predicted open reading frames (ORFs), or were localized in clearly predicted 5′ and 3′ untranslated regions (UTRs).

A covariance-model-based algorithm[13] was used to search all unassigned cDNA reads for both sequence and structural similarity to known sRNA families archived in the RNA families database Rfam[14]. Thirteen known sRNA families were captured in the environmental transcriptomes, representing only ~16% of the total reads detected by IGR mapping. The most abundant sRNAs belonged to ubiquitous or highly conserved sRNA families including transfer-messenger RNA (tmRNA), RNase P RNA, signal recognition particle RNA (SRP RNA) and 6S RNA (SsrS RNA; Supplementary Table 1). In addition, a number of known riboswitches (*cis*-acting regulatory elements that regulate gene expression in response to ligand binding[15]) were detected in lower abundance, including glycine, thiamine pyrophosphate, cobalamin and *S*-adenosyl methionine riboswitches (Supplementary Table 1). The apparent taxonomic origins of the most abundant known sRNAs revealed depth-specific variation that was generally, but not always, consistent with known microbial depth distributions[4] (Supplementary Fig. 2). For example, although SRP RNAs are abundant in our data sets, very few *Pelagibacter*-like SRP

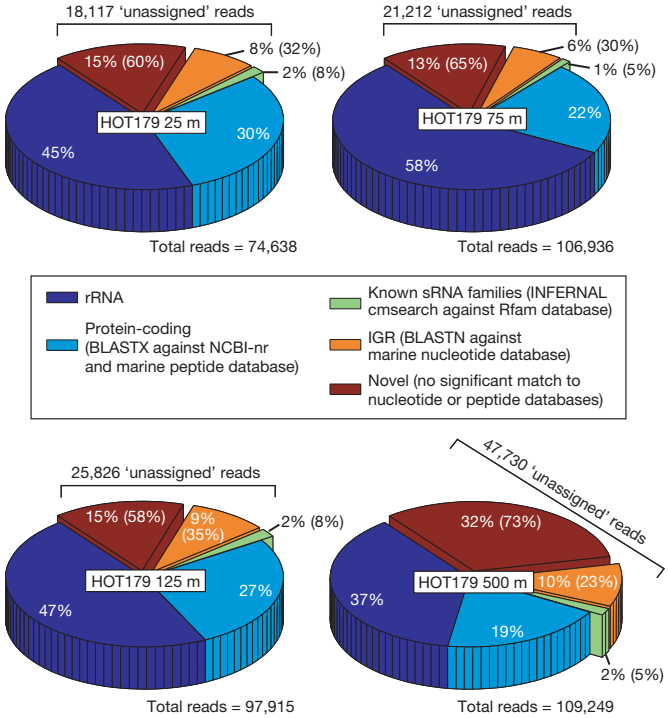


Figure 1 | Inventory of RNAs from each depth in the microbial metatranscriptomic datasets. The three offset slices represent reads that are not assigned to rRNA or known protein-coding genes, and are referred to as 'unassigned'. Numbers in parentheses represent the percentage of the total unassigned cDNA reads in each category.

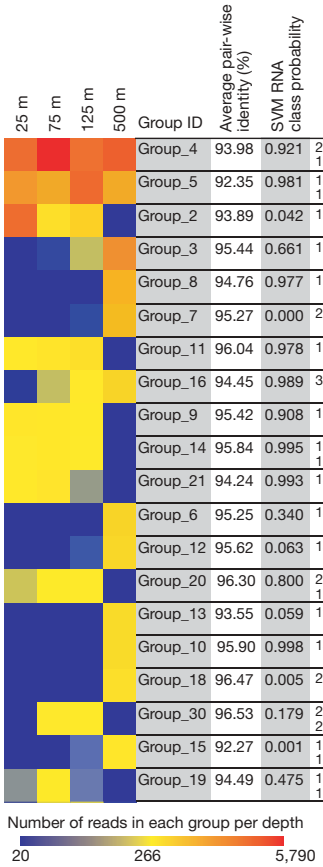
[1]Department of Civil and Environmental Engineering, and [2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

RNA reads were detected, indicating that SRP-dependent protein recognition and transport may not be a dominant form of protein translocation in oceanic *Pelagibacter* populations.

For better characterization of sRNAs in our data sets, including previously unknown sRNA families (referred to as putative sRNAs (psRNAs) hereafter), we pooled all cDNA reads from each sample and used a self-clustering approach to group homologous cDNA reads (see Methods). On the basis of observations from the IGR mapping (Supplementary Fig. 1), the self-clustering approach would help identify potential sRNAs because they are likely to span short genomic regions and exhibit high abundance (in many cases orders of magnitude higher than transcripts of protein-coding genes found in the same data sets). A total of 66 groups that comprised at least 100 overlapping cDNA reads were identified (Fig. 2 and Supplementary Table 2). For several of these groups, the abundance and depth-dependent distribution detected by means of cDNA pyrosequencing was confirmed using reverse transcription–quantitative polymerase chain reaction (RT–qPCR) analyses (Supplementary Fig. 3). Among the 66 groups, 9 were identified as belonging to Rfam sRNA families (Supplementary Table 2), and most of the remaining psRNA groups mapped to IGRs on metagenomic fragments derived from marine planktonic microorganisms.

Although they bear no resemblance to known peptide sequences, the psRNA groups could potentially represent mRNA degradation products or small unannotated protein-coding regions. We applied several criteria to help rule out these possibilities, including location within IGRs, psRNA length, lack of coding potential and conserved secondary structure. First, the psRNAs ranged in size between 100 and 500 nucleotides (Supplementary Fig. 4 and Supplementary Table 2), and tended to have an increased GC content when located within an AT-rich genome context[16] (Fig. 3a). Second, we systematically screened multiple sequence alignments of all 66 groups for coding potential, as indicated by three-base periodicity in the nucleotide substitution patterns[17] (Methods). Only sequences in group 92 were identified as possibly encoding proteins (Fig. 3b), and these were subsequently mapped to a specific hypothetical protein (NCBI accession number: ABZ07689) from a recently described uncultured marine crenarchaeote[18]. Third, the psRNA groups encompassed relatively divergent sequences that internally shared conserved secondary structures (for example, Fig. 3a, inset), indicating evolutionary coherence of functional roles and mechanisms. The alignment of full-length psRNA sequences revealed clear nucleotide co-variation that preserved base pairing in the consensus secondary structure (for example, Supplementary Fig. 5). In a specific example (group 5), although three divergent *Pelagibacter*-like psRNA sequences (one from 4,000 m depth[18] and two from surface waters[19]) shared pairwise nucleotide identities of only 78% to 87%, predicted secondary structures were nearly identical (Supplementary Fig. 6). Although computational analyses alone cannot be completely definitive, these combined criteria support our hypothesis that most of the psRNA groups that we identified represent authentic microbial sRNAs.

Many of the psRNAs identified here may be derived from as-yet uncharacterized microorganisms. For example, nine self-clustered psRNA groups shared no obvious homology with known nucleotide



Group ID	Average pair-wise identity (%)	SVM RNA class probability	Average length	Location	Flanking ORF annotations	Putative taxonomy of environmental fragments
Group_4	93.98	0.921	212 bp 173 bp	5′ 3′	COG0469: pyruvate kinase COG2234: predicted aminopeptidases	Bacteria; Proteobacteria
Group_5	92.35	0.981	148 bp 155 bp	5′ 3′	COG0350: methylated DNA-protein cysteine methyltransferase COG0482: tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_2	93.89	0.042	189 bp	3′	Hypothetical protein	Bacteria; Cyanobacteria; Prochlorales (MIT9215 strain specific)
Group_3	95.44	0.661	170 bp	3′	COG3202: ATP/ADP translocase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_8	94.76	0.977	157 bp	5′	Hypothetical protein	Bacteria; Proteobacteria; Gammaproteobacteria
Group_7	95.27	0.000	205 bp	5′	COG1254: acylphosphatases	Archaea; Crenarchaeota
Group_11	96.04	0.978	187 bp	5′	COG0077: prephenate dehydratase	Bacteria; Proteobacteria; Deltaproteobacteria/Gammaproteobacteria
Group_16	94.45	0.989	362 bp	3′	COG0114: fumarase	Bacteria; Spirochaetes
Group_9	95.42	0.908	191 bp	5′	COG0571: dsRNA-specific ribonuclease	Bacteria; Cyanobacteria; Prochlorales
Group_14	95.84	0.995	198 bp 158 bp	5′ 3′	COG0116: predicted N6-adenine-specific DNA methylase Hypothetical protein	Bacteria; Cyanobacteria; Prochlorales
Group_21	94.24	0.993	139 bp	5′	COG0180: tryptophanyl-tRNA synthetase	Bacteria; Proteobacteria; Gammaproteobacteria
Group_6	95.25	0.340	138 bp*	Unknown	No hits	Unassigned
Group_12	95.62	0.063	135 bp	Unknown	No hits	Unassigned
Group_20	96.30	0.800	201 bp 157 bp	Unknown 3′	COG0822: NifU homolog involved in Fe-S cluster formation COG1804: predicted acyl-CoA transferases/carnitine dehydratase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_13	93.55	0.059	194 bp	Unknown	No hits	Archaea; Crenarchaeota
Group_10	95.90	0.998	133 bp*	Unknown	No hits	Unassigned
Group_18	96.47	0.005	275 bp	3′	COG0531: amino acid transporters	Archaea; Euryarchaeota
Group_30	96.53	0.179	222 bp 220 bp	5′ 3′	COG0596: predicted hydrolases or acyltransferases Conserved hypothetical protein	Bacteria; Cyanobacteria; Prochlorales
Group_15	92.27	0.001	107 bp 101 bp	5′ 3′	COG1245: predicted ATPase, RNase L inhibitor (RLI) homologue No hits	Archaea; Crenarchaeota
Group_19	94.49	0.475	173 bp 173 bp	3′ 5′	No hits COG1629: outer membrane receptor proteins, mostly iron transport	Bacteria; Proteobacteria; Gammaproteobacteria

Figure 2 | Abundance, distribution and features of the top twenty most abundant sRNA and psRNA groups identified in the metatranscriptomic data. The twenty groups were ranked based on total abundance. Each group's depth distribution is shown in the left panel, with the number of reads in each data set indicated by colour, from high (red) to low (blue). Each group's proximity (5′ or 3′) to the nearest gene, annotation and putative taxonomy for that gene (where possible) are shown. The RNA-class probability values were generated with an SVM learning algorithm using RNAz[29]. Group 9 is comprised of *Prochlorococcus*-like RNase P RNAs. Group 21 sRNAs probably mediate regulation (via transcription attenuation) of tryptophanyl tRNA synthetase. Group 30 contains overlapping sRNAs *Yfr8* and *Yfr9* indentified in *Prochlorococcus* MED4 in ref. 8. Lengths of putative sRNAs with no homology with known nucleotide sequences (each marked with an asterisk) were predicted through assembly of cDNAs from each group (average contig size, see Methods). A complete list of sRNA and psRNA groups containing >100 cDNA reads is provided in Supplementary Table 2.

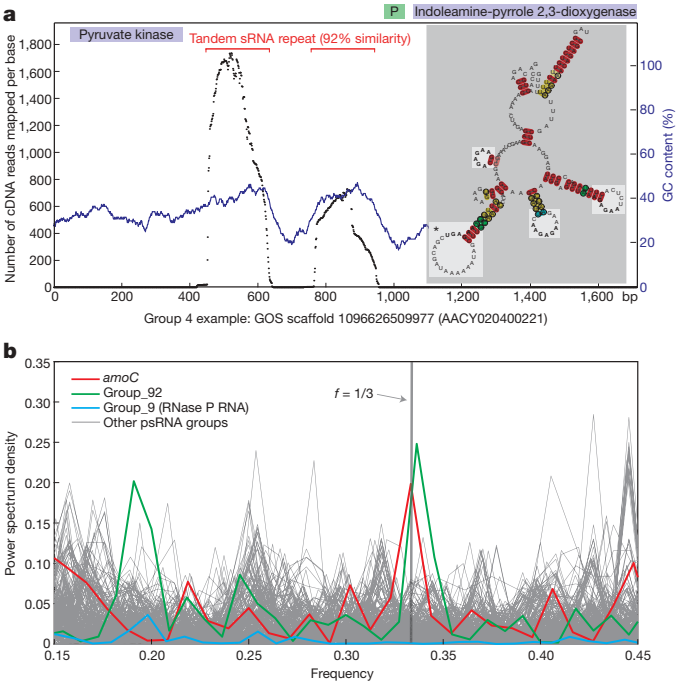


Figure 3 | Characteristics of psRNA groups consistent with known sRNAs. **a**, Genomic context and features of the most abundant psRNA group, group 4, mapped onto a Gammaproteobacteria-like contig from the Global Ocean Sampling (GOS) database. Sequence coverage (black dots, left axis) and reference GC content (blue dots, right axis) are shown. Gene annotations are indicated along the top of the panel (annotated ORFs shown outside and inside the panel are on the forward and reverse strand, respectively; P represents promoter). In the predicted structure (inset), loops containing conserved sequence motifs (in bold letters) are highlighted, and the loop marked with an asterisk contains sequences predicted to interact with 5′ translation start site of a flanking gene. **b**, Three-base periodicity analysis of multiple sequence alignments for the 66 self-clustered groups. A significant peak in the power spectrum density at the frequency of 1/3 indicates three-base periodicity in the nucleotide substitution patterns, indicating protein-coding potential[17]. See Methods.

sequences (for example, groups 6 and 10), and seem to represent completely new sRNA families. Most of these were found only in the 500 m sample (Fig. 2). The remaining psRNA groups mapped to IGRs on genomic and metagenomic sequences derived from planktonic marine microbes. Although identifying sRNA regulatory functions and their target genes is a major challenge even for model microorganisms[20], the conserved genomic context of these psRNAs has potential to provide insight into their functional roles[21,22]. The most predominant gene families flanking these psRNA groups included transporter genes involved in nutrient acquisition (inorganic nitrogen, amino acids, iron and carbohydrates) and genes involved in energy production and conversion (Supplementary Table 2). These results highlight the potential importance of sRNA regulation of nutrient acquisition and energy metabolism in free-living planktonic microbial communities.

The most populated psRNA cluster, group 4, appeared to be involved in the regulation of central carbon metabolism and energy production in Proteobacteria (predominantly Gammaproteobacteria). The psRNAs from this group were flanked by genes involved in pyruvate metabolism (for example, pyruvate kinase and malate synthase), glucose transport (sodium glucose symporter) and nitrogen acquisition (ammonia permease and aminopeptidase; Fig. 2 and Supplementary Table 2). In several cases, group 4 psRNAs occurred in tandem copies within the same IGR (Fig. 3a). Small RNAs that display stable secondary structure typically mediate regulation using sequences in loop domains to interact with specific target sequences[3,23]. Consistent with this mechanism, a conserved six-nucleotide sequence motif (AAGAGN) appeared in multiple loops within predicted hairpin structures for group 4 (Fig. 3a, inset). The six-nucleotide sequence AAGAGA was previously verified as a ribosomal binding site[24], and indicates that group 4 psRNAs may have a regulatory role at the translational level. Indeed, sequences in one of the loop domains of the consensus structure (Fig. 3a, inset) have potential to interact (by base pairing across 32 bp) with the flanking pyruvate kinase gene near the 5′ translation initiation site.

In contrast to the broad taxonomic affiliations of group 4 psRNAs, the other highly abundant psRNA group, group 5, appeared almost exclusively on *Pelagibacter*-like genomic fragments recovered from both open ocean surface waters[19] and abyssal (4,000 m) depth[18], but did not map to the genomes of currently cultivated *Pelagibacter* strains (Fig. 2 and Supplementary Table 2). Group 5 psRNAs mapped onto 203 different metagenomic fragments, predominantly in the 5′ UTR of 6-*O*-methylguanine DNA methyltransferase (6-*O*-MGMT, COG0350; involved in DNA repair) and the 3′ UTR of tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (*trmU*, COG0482; involved in tRNA modification). A predicted promoter and Rho-independent terminator flanked group 5 psRNAs upstream of 6-*O*-MGMT, and attenuator/riboswitch characteristics were identifiable in the 5′ UTR by secondary structure prediction (Supplementary Fig. 6). Indeed, the presence of riboswitch-like elements upstream of 6-*O*-MGMT genes was previously predicted by comparing 223 complete bacterial genomes[25].

Unlike group 4 and 5 psRNAs, the remaining self-clustered sRNA and psRNA groups showed depth-variable distributions (Fig. 2). Group 7 psRNAs were enriched at 500 m and were highly conserved in marine crenarchaeal genomes. Similarly, cyanobacteria-like psRNAs were enriched in the photic zone (for example, groups 2, 30, 48 and 17; Supplementary Table 2). One of these groups (group 30) includes two experimentally validated sRNAs (*Yfr8* and *Yfr9*), which were found antisense to one another and were hypothesized to be involved in a toxin–antitoxin system in *Prochlorococcus marinus* MED4 (ref. 8). Intriguingly, a few *Prochlorococcus*-like psRNA groups mapped to some but not all coexisting members of the *Prochlorococcus* population, indicating that such sRNAs may provide niche-specific regulation. Group 2 psRNAs, for example, were detected only in the genome of *P. marinus* strain MIT9215 and in a highly similar genomic fragment from the environment (NCBI accession number: DQ366713). Group 2 psRNAs are located in a hyper-variable region adjacent to phosphate transporter genes, and share a 14-bp exact match with the 5′ translation initiation site of the phosphate ABC transporter gene (*pstC*). In *Prochlorococcus* strains lacking the phosphate regulon two-component response regulator (*phoB*) and signalling kinase (*phoR*)[26], such as MIT9215, it is possible that sRNAs represent an alternative mechanism for regulating phosphorus assimilation.

To examine sRNA representation in specific abundant microbial groups, we aligned the psRNA reads to the genome of an abundant planktonic bacterium, *Candidatus Pelagibacter ubique* HTCC7211. Eleven IGRs on the *P. ubique* HTCC7211 genome coincided with the psRNAs identified in our samples (Fig. 4), 6 of which were also independently predicted to be sRNA-containing IGRs (support vector machine, SVM, RNA-class probability >0.9) by comparative analysis of three *P. ubique* genomes (Methods and Supplementary Table 3). Genes flanking these expressed psRNAs included DNA-directed DNA polymerase gamma/tau subunit (*dnaX*), *carD*-like transcriptional regulator family and alternative thymidylate synthase (Supplementary Table 3). Notably, covariance-model-based searches identified cDNAs mapping to glycine riboswitch motifs in two *Pelagibacter* IGRs (Fig. 4 and Supplementary Table 3). Recently, it was experimentally verified that *P. ubique* HTCC1062 uses one of these two glycine riboswitches to sense the intracellular glycine level and to regulate its carbon usage for biosynthesis and energy[27].

The diversity and abundance of sRNAs in microbial metatranscriptomic data sets indicates that natural microbial assemblages use a wide variety of sRNAs for regulating gene expression in response to variable environmental conditions. The data and analyses

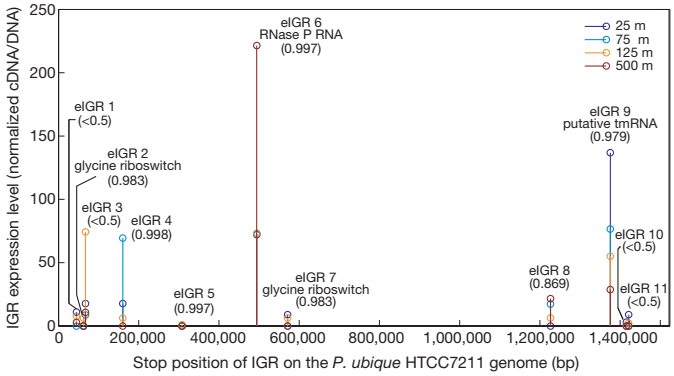


Figure 4 | Normalized cDNA/DNA ratios of expressed IGRs (eIGRs) on the *P. ubique* HTCC7211 genome at all four depths. Because a manually curated HTCC7211 genome annotation is not yet publicly available, the genomic regions that recruited psRNAs were manually inspected and confirmed as IGRs. The values in the parentheses are RNA-class probability values generated with a SVM learning algorithm using RNAz[29].

described here provide a culture-independent tool to expand our knowledge of the sequence motifs, structural diversity and genomic distributions of microbial sRNAs that are expressed under specific environmental conditions. Although the exact regulatory functions of many of the psRNAs remain to be experimentally verified, their *in situ* expression, structural features and genomic context all provide a solid foundation for future studies. These data, in conjunction with metatranscriptomic field experiments linking environmental variation with changes in RNA pools, have potential to provide new insights into environmental sensing and response in natural microbial communities.

METHODS SUMMARY

Bacterioplankton samples were collected from the HOT Station ALOHA (22° 45′ N, 158° W) in March 2006 at four different depths (25 m, 75 m, 125 m and 500 m), and immediately frozen and stored at −80 °C until processing. Nucleic acid extraction, RNA amplification, cDNA synthesis and pyrosequencing were performed as previously described[1]. Ribosomal RNA sequences were identified by querying against a comprehensive rRNA database using BLASTN, and were excluded from the subsequent sRNA analysis. Protein-coding genes were recognized by querying with BLASTX against published peptide databases as well as a custom marine-specific peptide database (Methods). A covariance-model-based program (INFERNAL)[13] was used to search for known sRNA elements in the data sets. The self-clustering approach (see Methods) to identify abundant psRNAs in the environment was based on sRNA reads spanning across a short genomic region in high abundance. Self-clustered groups that contained more than 100 cDNA reads were further characterized in detail, including secondary structure prediction using RNAalifold[28], coding potential evaluation, genomic context examination and sRNA-class probability calculation using RNAz[29] (see Methods). The genome sequences of an oceanic *Pelagibacter* strain (HTCC7211) were used to recruit psRNA reads to examine possible regulatory sRNAs related to oceanic *Pelagibacter* populations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 October 2008; accepted 9 April 2009.

1. Frias-Lopez, J. *et al.* Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA* **105,** 3805–3810 (2008).
2. Gilbert, J. A. *et al.* Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* **3,** e3042 (2008).
3. Storz, G. & Haas, D. A guide to small RNAs in microorganisms. *Curr. Opin. Microbiol.* **10,** 93–95 (2007).
4. DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311,** 496–503 (2006).
5. Gottesman, S. Stealth regulation: biological circuits with small RNA switches. *Genes Dev.* **16,** 2829–2842 (2002).
6. Lenz, D. H. *et al.* The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118,** 69–82 (2004).
7. Duehring, U., Axmann, I. M., Hess, W. R. & Wilde, A. An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc. Natl Acad. Sci. USA* **103,** 7054–7058 (2006).
8. Steglich, C. *et al.* The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet.* **4,** e1000173 (2008).
9. Vogel, J. *et al.* RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31,** 6435–6443 (2003).
10. Silvaggi, J. M., Perkins, J. B. & Losick, R. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *J. Bacteriol.* **188,** 532–541 (2006).
11. Karl, D. M. & Lukas, R. The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep-Sea Res. II* **43,** 129–156 (1996).
12. Kawano, M., Reynolds, A. A., Miranda-Rios, J. & Storz, G. Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **33,** 1040–1050 (2005).
13. Eddy, S. *INFERNAL User's Guide, Version 0. 72* ⟨http://infernal.janelia.org/⟩ (2007).
14. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33,** D121–D124 (2005).
15. Brantl, S. Bacterial gene regulation: from transcription attenuation to riboswitches and ribozymes. *Trends Microbiol.* **12,** 473–475 (2004).
16. Schattner, P. Searching for RNA genes using base-composition statistics. *Nucleic Acids Res.* **30,** 2076–2082 (2002).
17. Ré, M. & Pavesi, G. in *Applications of Fuzzy Sets Theory* (eds Masulli, F., Mitra, S. & Pasi, G.) 544–550 (Springer, 2007).
18. Konstantinidis, K. T. & DeLong, E. F. Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J.* **2,** 1052–1065 (2008).
19. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** 398–431 (2007).
20. Vogel, J. & Wagner, E. G. H. Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.* **10,** 262–270 (2007).
21. Hershberg, R., Altuvia, S. & Margalit, H. A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.* **31,** 1813–1820 (2003).
22. Yao, Z. *et al.* A computational pipeline for high-throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. *PLoS Comp. Biol.* **3,** 1212–1223 (2007).
23. Trotochaud, A. E. & Wassarman, K. M. A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Struct. Mol. Biol.* **12,** 313–319 (2005).
24. Bruttin, A. & Brüssow, H. Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. *Virology* **219,** 96–104 (1996).
25. Abreu-Goodger, C. & Merino, E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* **33,** W690–W692 (2005).
26. Martiny, A. C., Coleman, M. L. & Chisholm, S. W. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc. Natl Acad. Sci. USA* **103,** 12552–12557 (2006).
27. Tripp, H. J. *et al.* Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environ. Microbiol.* **11,** 230–238 (2008).
28. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
29. Washietl, S., Hofacker, I. L. & Stadler, P. F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA* **102,** 2454–2459 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the University of Hawaii HOT team, and the captain and crew of the RV *Kilo Moana* for their expert assistance at sea. We also thank S. Schuster for collaboration and advice on pyrosequencing, J. Eppley for help with computational analyses and discussion, and J. Maresca, A. Martinez, J. McCarren and V. Rich for their comments on this manuscript. We thank S. Giovannoni, J. Tripp and M. Schwalbach for sharing their in press manuscript on *Pelagibacter* riboswitches, and S. Giovannoni, U. Stingl, the J. Craig Venter Institute and the Gordon and Betty Moore Foundation for the genome sequence of *Pelagibacter* strain HTCC7211. This work was supported by the Gordon and Betty Moore Foundation, National Science Foundation Microbial Observatory Award MCB-0348001, the Department of Energy Genomics GTL Program, the Department of Energy Microbial Genomics Program, and an NSF Science and Technology award, C-MORE. This article is a contribution from the NSF Science and Technology Center for Microbial Oceanography: Research and Education (C-MORE).

Author Contributions E.F.D. conceived and directed the research, coordinated the sequencing effort and collected the samples. Y.S. prepared samples for sequencing, and made the initial observation of sRNA sequences. E.F.D., Y.S. and G.W.T. developed the concept of the paper together. Y.S. and G.W.T. performed the data analysis. Y.S. wrote the first draft of the paper, which was completed by G.W.T. and E.F.D. together.

Author Information The sequences reported here have been deposited in GenBank under accession numbers SRA007802.3, SRA000263, SRA007804.3 and SRA007806.3 corresponding to cDNA sequences, and SRA007801.5, SRA000262, SRA007803.3 and SRA007805.4 corresponding to DNA sequences, for 25 m, 75 m, 125 m and 500 m samples, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.F.D. (delong@mit.edu).

METHODS

Sample collection and RNA/DNA extraction. Bacterioplankton samples from the photic zone (25 m, 75 m, 125 m) and the mesopelagic zone (500 m) were collected from the HOT Station ALOHA site in March 2006, as described previously[1]. In brief, four replicate 1-l seawater samples were prefiltered through 1.6-mm GF/A filters (Whatman) and then filtered onto 0.22-μm Durapore filters (25 mm diameter, Millipore) using a four-head peristaltic pump system. Each Durapore filter was immediately transferred to screw-cap tubes containing 1 ml of RNAlater (Ambion Inc.), and frozen at −80 °C aboard the RV *Kilo Moana*. Samples were transported frozen to the laboratory in a dry shipper and stored at −80 °C until RNA extraction. Total sampling time, from arrival on deck to fixation in RNAlater, was less than 20 min.

Total RNA was extracted as previously described[1], using the mirVana RNA isolation kit (Ambion), with several modifications as follows. Samples were thawed on ice, and then 1-ml RNAlater was loaded onto two Microcon YM-50 columns (Millipore) to concentrate and desalt each sample. The resulting 50 μl of RNAlater was added back to the sample tubes, and total RNA extraction was performed following the mirVana manual. Genomic DNA was removed using a Turbo DNA-free kit (Ambion). Finally, extracted RNA (DNase-treated) from four replicate filters was combined, purified and concentrated using the MinElute PCR Purification Kit (Qiagen).

Bacterioplankton sampling for DNA extraction was performed as previously described[1].

Complementary DNA synthesis and sequencing. The synthesis of microbial community cDNA from small amounts of mixed-population microbial RNA was performed as previously described[1]. In brief, nanogram quantities of total RNA were polyadenylated using *Escherichia coli* poly(A) polymerase I (E-PAP)[30]. First-strand cDNA was then synthesized using ArrayScript (Ambion) with an oligo(dT) primer containing a T7 promoter sequence and a restriction enzyme (BpmI) recognition site sequence, followed by the second-strand cDNA synthesis[1]. The double-stranded cDNA templates were transcribed *in vitro* using T7 RNA polymerase at 37 °C for 6 h[31], yielding a large amount of antisense RNA. The SuperScript double-stranded cDNA synthesis kit (Invitrogen) was used to convert antisense RNA to microgram quantities of cDNA, which was then digested with BmpI to remove poly(A) tails. Purified cDNA was then directly sequenced by pyrosequencing[32].

Removal of low-quality and rRNA GS20 cDNA sequences. Low-quality cDNA reads were removed as previously described[1].

Reads encoding rRNA were identified and removed from the cDNA data sets by comparing them to a combined 5S, 16S, 18S, 23S and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (http://www.arb-silva.de). BLASTN[33] matches with bit score ≥50 were considered significant and deemed rRNA sequences. In test simulations, this bit score cutoff resulted in <1.7% false positives against a database of all non-rRNA microbial genes from available microbial genomes.

Identification of protein-coding genes. Protein-coding cDNA reads were identified by translating nucleotide sequences in all 6 frames and comparing each to Global Ocean Sampling peptides, the NCBI-nr protein database and a custom peptide database using BLASTX[33]. The custom peptide database contained marine-specific ORF sequences predicted from four sources: the Moore Microbial Genome Project genomes (http://www.moore.org/microgenome/strain-list.aspx), large genome fragments (~40 kb) from a variety of marine habitats (Rich *et al.*, in preparation), and both fosmid end sequences and shotgun library sequences generated from depth profile bacterioplankton samples collected in multiple HOT cruises (E.F.D. *et al.*, in preparation). Unpublished databases are available on request.

After rRNA sequences were removed, each cDNA data set contained between 40,000 and 70,000 pyrosequence reads. Of these cDNA reads, a large fraction (~50% of those from photic-zone samples; ~70% from the mesopelagic sample) showed no significant homology to either the non-redundant peptide database from NCBI or marine microbial peptide sequences, using the bit score of 40 that has been previously validated as a cutoff for calling homology in short pyrosequencing reads[1].

Assignment of cDNA reads to known non-coding RNA families. We searched the Rfam database[14] to investigate the representation and diversity of known sRNA families in our data sets. Rfam is a collection of non-coding RNA families, represented by multiple sequence alignments and covariance models, including those from 400 complete genomes including 233 bacterial and 24 archaeal genomes (June 2008 version). The INFERNAL program (http://infernal.janelia.org/) was used to search for RNA structure and sequence similarities based on covariance models (also called profile stochastic context-free grammars)[34]. The reference database was a collection of covariance models for all non-coding RNA families downloaded from the Rfam (version 8.1) ftp site (http://www.sanger.ac.uk/Software/Rfam/ftp.shtml). A perl wrapper named Rfamscan.pl (http://www.sanger.ac.uk/Software/Rfam/help/software.shtml), written by Sam Griffiths-Jones, was used to run batch queries (>200,000 cDNA reads) on a local machine.

To test the specificity and sensitivity of the INFERNAL Rfam-seeded search of our cDNA reads, two data sets were created from the *E. coli* strain K12 substrain MG1655, in which sRNAs have been well defined[35]. The two test data sets were protein-coding sequences and known sRNA sequences, each with the same length distributions as our cDNA data set (that is, 206,418 sequence fragments with mean sequence length 97 bp). The INFERNAL Rfam-seeded search of the *E. coli* MG1655 protein-coding test data set yielded no significant hits, indicating high specificity and a false-positive rate below detection. However, the INFERNAL Rfam-seeded search did not identify all *E. coli* MG1655 sRNA fragments, probably owing to the short lengths of the query sRNA fragments. To compensate for the decreased search sensitivity due to shorter read length, we queried all cDNA reads against all full-length sRNA sequences in the Rfam database by BLASTN. Reads that did not meet the default cutoffs defined by Rfamscan, but shared good homology with Rfam member sequences by BLASTN (alignment length ≥90% of sequence length; sequence identity ≥85%), were also assigned to the corresponding sRNA families.

Putative taxonomic assignment of cDNA reads in known sRNA families. Potential taxonomic origins of the known sRNAs were investigated by searching against NCBI-nt (4 July 2008) using BLASTN (word size of 7, default e-value cutoff, low complexity filter off, and the ten best hits retained). The BLASTN results were then parsed using MEGAN[36] using default parameters—that is, the congruent taxonomy of the hits that were within 10% below the best hit was assigned to the cDNA read.

Self-clustering approach to identify sRNA and psRNA groups. A self-clustering approach allowed related cDNA reads to form distinct groups that could be separated from other transcripts based on sequence similarity and overall abundance. Combined cDNA reads (206,418 reads after the removal of rRNAs) from all 4 depths were locally aligned to each other (that is, all sequences served both as queries and subjects) using BLASTN with the following settings different from default: $W = 7$, $F =$ F, $m = 8$, $v = 206418$, $b = 206418$, $e = 1 \times 10^{-5}$. A perl script was used to group similar cDNA reads based on the BLASTN output. In brief, for each cDNA query, all matches that met a minimum cutoff of 85% sequence identity over 90% average sequence length were considered significant and stored into a hash. The hash then was ranked on the basis of the number of matches stored for each hash key (query). The cDNA read with the most matches served as a seed sequence of the first cluster. After all matches of the seed sequence were recruited, the script looped over each one of the matches and gathered all subsequent matches until the chain disconnected and a new cluster started to form.

The self-clustering approach was successful in identifying a number of highly abundant psRNA groups. These psRNAs were clearly distinct from protein-coding clusters as they were found in much higher copy number than most mRNAs, and the typical length of psRNAs was ~100–500 nucleotides. The sequence identity cutoff (85%) was chosen because it allowed known RNase P RNAs from closely related microbial populations (for example, all *Prochlorococcus* RNase P RNAs) to form a distinct sequence group. However, because sRNA species by nature differ in their primary sequence divergence, clustering based on one sequence identity cutoff inevitably yields psRNA groups with different within-group diversity, which either represent homologues from closely related microbial populations or highly conserved elements from diverse microbial taxa.

Systematic screening for coding potentials of the self-clustered groups. We identified a total of 66 groups that contained more than 100 cDNA reads (a file named 'H179_sRNA_groups.tgz', containing all sequences from these 66 groups, and a file named 'H179_sRNA_groups_CLUSTAL.tgz', containing multiple sequence alignments of subsets of sequences from these 66 groups, can be downloaded from http://web.mit.edu/ymshi/Public/). To assess the possibility that some groups represent unannotated small proteins, we systematically screened multiple sequence alignments of these 66 groups for coding potentials based on three-base periodicity in nucleotide substitution patterns. The rationale of detecting three-base periodicity in coding regions is that codons encoding the same amino acid often differ only in a single nucleotide located in the third position of the codon. As a direct consequence, in coding sequences under selective evolutionary pressure, substitutions are more often tolerated if they occur at the third position of codons. Therefore, if aligned sequences are protein-coding, the spectral signal of the mismatches along the alignment is expected to be maximal at frequency 1/3 (three-base periodicity)[17].

We generated a pipeline for multiple sequence alignment, nucleotide diversity calculation (conversion of DNA sequence alignments to numerical sequences) and Fourier transform and power spectrum analysis of the numerical sequences for all 66 groups (including known sRNAs and psRNAs). Specifically, 100

sequences were randomly sampled from a subset of overlapping sequences in each group, and aligned using MUSCLE 3.6 (ref. 37). The random sampling and alignment was repeated multiple times proportional to the number of sequences in the group. For each alignment, average nucleotide diversity was calculated for each column of the alignment as following:

$$D_{\text{average}} = \sum D_{\text{pair-wise}} \Big/ N(N-1)/2$$

where D_{average} represents average nucleotide diversity, $D_{\text{pair-wise}}$ represents pair-wise nucleotide diversity (a pair of identical nucleotides was given a value of 0, and a pair of different nucleotides was given a value of 1) and $N(N-1)/2$ represents the total number of pairs in the column of the alignment. Owing to high insertion/deletion error rate of pyrosequencing[32], any alignment column where greater than 75% of sequences had a gap resulted in that column being ignored in the subsequent calculation. After the multiple sequence alignments were converted to numerical sequences, a Fourier transform and power spectrum analysis[38] of the numerical sequences were performed using MATLAB (http://www.mathworks.com/) to find significant frequencies of periodicity.

RT–qPCR analysis of psRNA group 7 and sRNA group 9. The apparent abundance and depth-dependant distribution of group 7 and group 9 in our metatranscriptomic data sets were validated using RT–qPCR. Owing to lack of absolute quantification standards for these groups, we calculated their relative abundance to the crenarchaeal ammonia monooxygenase subunit A (*amoA*) transcript in the 500 m sample. Primers for these groups were designed using the Invitrogen web-based OligoPefect primer designer. The primer sequences are: G7_Primer1 (5′-AGCTCTGCTGGTTCYAGACT-3′) and G7_Primer2 (5′-TCGAACATTCACGCTTCCT-3′); G9_Primer1 (5′-TAAGCCGGGTTCTG TTCATC-3′) and G9_Primer2 (5′-GCCGCTTGAGACTGTGAAGT-3′). The primer set for the crenarchaeal *amoA* transcript was the same as previously published[39]: CrenAmoAQ-F (5′-GCARGTMGGWAARTTCTAYAA-3′) and CrenAmoAModR (5′-AAGCGGCCATCCATCTGTA-3′). All primers were blasted against NCBI-nt database to avoid potential matches to unwanted regions.

Possible traces of DNA were removed from all RNA samples using the Turbo DNA-free kit (Ambion) following the manufacturer's instructions. For each reverse transcription reaction, 1 μl of RNA (4–7.5 ng) was reverse transcribed using gene-specific primer and Superscript III reverse transcriptase (Invitrogen). Reverse transcription was performed at 50 °C for 50 min, after an initial incubation step of 5 min at 65 °C. The reverse transcription reactions were terminated at 85 °C for 5 min, and 1 μl RNase H was added to each reverse transcription reaction, followed by incubation at 37 °C for 20 min. Subsequently, SYBR Green qPCR reactions were performed on LC480 (Roche Applied Science) using the specific primer set for each gene of interest. We used the $2^{-\Delta\Delta CT}$ method[40] to compare the relative abundance of group 7 and group 9 transcripts in all 4 samples (25 m, 75 m, 125 m and 500 m) to the crenarcheal *amoA* transcript in the 500 m sample.

Characterizing psRNA groups. The psRNA groups were further characterized to determine the approximate psRNA length, proximity to (5′ or 3′ or unknown (when the psRNA is not flanked by one ORF on each side)) and annotation of nearest flanking ORF on available genome/metagenome fragments, putative taxonomy and SVM-based RNA class probability. Pooled cDNA reads (not including rRNA reads) from each transcriptomic data set were queried against a custom database of nucleotide sequences from available genome and metagenomic projects (see above) using BLASTN. Metagenomic fragments in this database were run through Metagene[41] to identify predicted ORFs (coding) and intergenic (non-coding) regions.

Using the BLASTN and Metagene results, cDNA reads were mapped to each genome/metagenome fragment based on sequence similarity (≥85% identity over 90% of the read length), which could be used to calculate coverage values for each coding and intergenic region on each genomic/metagenomic fragment. Two groups were identified as highly expressed protein-coding genes (group 35, ammonia monooxygenase subunit C; and group 42, ammonia permease) and were excluded from further analyses. In most cases, reads belonging to putative sRNA groups mapped with high coverage to IGRs on genomic/metagenomic fragments. In these cases, we estimated the size of psRNAs in each group by defining the psRNAs as the sequence region in intergenic space having minimum sequence coverage of greater than ten times. In addition, it was also possible to determine the location of these psRNAs with respect to coding sequences. psRNAs were labelled as either 3′ or 5′ based on their position relative to the nearest flanking gene. Functional annotation for each of the genes flanking psRNA groups was obtained by comparing the amino acid sequences against the KEGG[42], COG[43] and the NCBI-nr databases from NCBI using BLASTP. Putative taxonomic origins of each fragment were assigned based on the NCBI taxonomy of matches in the NCBI-nr database.

Only 9 psRNA groups had no homology to sequences in the currently available database. To estimate the size of each of these psRNA groups, reads from each were assembled using PHRAP (−minmatch 15, −minscore 20, revise_greedy) and the average length of contigs (<10 contigs) formed used to infer sequence space spanned by the sRNA group.

To calculate the RNA class probability for each group, the first twenty cDNA reads recruited to each psRNA group were extracted from the data set and placed in the same sequence orientation. Multiple sequence alignments were performed using MUSCLE 3.6 (ref. 37). The sequence alignment for each psRNA groups (CLUSTALW format) was then used to predict consensus structure and the thermodynamic stability using RNAz[29], and an RNA-class probability was calculated based on the SVM regression analysis.

Secondary structure prediction. The minimum free energy structure was predicted based on the multiple sequence alignment of full-length psRNA sequences extracted from metagenomic sequence reads. The RNAalifold program from the Vienna RNA package[28,44] was used to produce consensus secondary structure and sequence alignment colour-coded based on nucleotide variations. The colour hue indicates how many of the six possible types of base-pairs (GC, CG, AU, UA, GU, UG) occur in at least one of the sequences. Pairs without sequence covariation are shown in red. Ochre, green, turquoise, blue and violet mark pairs that occur in two, three, four, five and six types of pairs, respectively. Pale colours mark pairs that cannot be formed by all sequences (that is, inconsistent base changes occur in some sequences). Attenuator-like structure was predicted using RibEx program[25].

Mapping cDNA reads to the genome of *P. ubique* HTCC7211. *Candidatus Pelagibacter ubique* HTCC7211 genome sequences were downloaded from the Moore Microbial Genome Project (http://www.moore.org/microgenome/strain-list.aspx). Based on the genome annotations, all IGR sequences greater than 50 bp (excluding rRNA and tRNA) were extracted and used to create BLASTN database. Both DNA and cDNA reads from each sample were then queried (BLASTN) against the database and parsed using same criteria as above (alignment length ≥90% of sequence length; identity ≥85%). For each IGR an expression ratio was calculated as the percentage of cDNA reads assigned to the IGR, relative to that in the DNA library. If there were cDNA hits but no DNA hits, the number of DNA hits was considered to be 1. This normalization compensates for the IGR length differences, and differences in DNA and cDNA library sizes.

Prediction of sRNA-containing IGRs in *Pelagibacter* genomes. Three *Pelagibacter* genomes (*Pelagibacter ubique* HTCC1062, HTCC1002 and HTCC7211) were used in the comparative genome analysis to predict possible sRNAs in the IGRs based on conserved secondary structure among closely related genomes[45]. A total of 1,113 IGRs were extracted from these three genomes (again only IGRs ≥50 bp and excluding tRNAs and rRNAs), and locally aligned to pooled ORFs and IGRs (5,398) from the three genomes using BLASTN with the following settings changed from default: $W = 7$, $F =$ F, $v = 5398$, $b = 5398$. ORFs were included so that *cis*-acting regulatory elements of mRNA were also examined. A total of 1,848 IGR sequences were extracted from all the high-scoring segment pairs with bit scores greater than 50, using Bioperl[46]. Self-clustering of this subset of *Pelagibacter* IGR sequences was then performed, as described above. Sequences in each cluster were aligned using MUSCLE 3.6 (ref. 37) and the alignments were scored for their secondary structure conservation and thermodynamic stability using RNAz 1.0 (ref. 29). SVM-based RNA-class probability values from the RNAz pipeline were gathered for each cluster and ranked from high to low.

30. Wendisch, V. F. *et al.* Isolation of *Escherichia coli* mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. *Anal. Biochem.* **290,** 205–213 (2001).
31. Vangelder, R. N. *et al.* Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl Acad. Sci. USA* **87,** 1663–1667 (1990).
32. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
33. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
34. Eddy, S. R. & Durbin, R. RNA sequence-analysis using covariance-models. *Nucleic Acids Res.* **22,** 2079–2088 (1994).
35. Rudd, K. E. EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.* **28,** 60–64 (2000).
36. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17,** 377–386 (2007).
37. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

38. Holste, D., Weiss, O., Grosse, I. & Herzel, H. Are noncoding sequences of *Rickettsia prowazekii* remnants of "neutralized" genes? *J. Mol. Evol.* **51,** 353–362 (2000).
39. Mincer, T. J. *et al.* Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific subtropical gyre. *Environ. Microbiol.* **9,** 1162–1175 (2007).
40. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-[Delta][Delta]CT Method. *Methods* **25,** 402–408 (2001).
41. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34,** 5623–5630 (2006).
42. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
43. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28,** 33–36 (2000).
44. Hofacker, I. L., Fekete, M. & Stadler, P. F. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* **319,** 1059–1066 (2002).
45. Axmann, I. M. *et al.* Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol.* **6,** R73 (2005).
46. Jason, S. & Ewan, B. The Bioperl project: motivation and usage. *SIGBIO Newsl.* **20,** 13–14 (2000).

LETTERS

Discovery of dual function acridones as a new antimalarial chemotype

Jane X. Kelly[1,2], Martin J. Smilkstein[1,2,3]†, Reto Brun[4], Sergio Wittlin[4], Roland A. Cooper[5], Kristin D. Lane[5]†, Aaron Janowsky[1,3], Robert A. Johnson[1,3], Rozalia A. Dodean[1,2], Rolf Winter[1,2], David J. Hinrichs[1,3] & Michael K. Riscoe[1,2,3]

Preventing and delaying the emergence of drug resistance is an essential goal of antimalarial drug development. Monotherapy and highly mutable drug targets have each facilitated resistance, and both are undesirable in effective long-term strategies against multi-drug-resistant malaria. Haem remains an immutable and vulnerable target, because it is not parasite-encoded and its detoxification during haemoglobin degradation, critical to parasite survival, can be subverted by drug–haem interaction as in the case of quinolines and many other drugs[1–5]. Here we describe a new antimalarial chemotype that combines the haem-targeting character of acridones, together with a chemosensitizing component that counteracts resistance to quinoline antimalarial drugs. Beyond the essential intrinsic characteristics common to deserving candidate antimalarials (high potency *in vitro* against pan-sensitive and multi-drug-resistant *Plasmodium falciparum*, efficacy and safety *in vivo* after oral administration, inexpensive synthesis and favourable physicochemical properties), our initial lead, T3.5 (3-chloro-6-(2-diethylamino-ethoxy)-10-(2-diethylamino-ethyl)-acridone), demonstrates unique synergistic properties. In addition to 'verapamil-like' chemosensitization to chloroquine and amodiaquine against quinoline-resistant parasites, T3.5 also results in an apparently mechanistically distinct synergism with quinine and with piperaquine. This synergy, evident in both quinoline-sensitive and quinoline-resistant parasites, has been demonstrated both *in vitro* and *in vivo*. In summary, this innovative acridone design merges intrinsic potency and resistance-counteracting functions in one molecule, and represents a new strategy to expand, enhance and sustain effective antimalarial drug combinations.

While feeding in the host red blood cell, malaria parasites ingest and degrade vast amounts of haemoglobin as a source of amino acids, consequently releasing toxic free haem as a by-product[6]. The parasites protect themselves from the toxic insult by converting haem into an insoluble crystalline material termed 'haemozoin' within the acidic digestive vacuole[7,8]. Because both haemoglobin degradation and haem detoxification are essential for parasite survival, these processes are important targets for antimalarial drug development[2]. The process of haem detoxification is widely believed to be the primary target of quinoline antimalarials (such as chloroquine and quinine), and it remains one of the most attractive and durable drug development targets[1,3–5], particularly because the complexity of the digestive vacuole environment and the immutable nature of the haem molecule have probably delayed the development of quinoline-resistant malaria for decades (chloroquine) or centuries (quinine) in the past[9].

The now-evident resistance to chloroquine is directly associated with mutations in the gene encoding the digestive vacuole membrane protein *Plasmodium falciparum* chloroquine resistance transporter (PfCRT), which results in reduced drug concentration at the target without altering the haem target itself[10–13]. In this case, in contrast to drug resistance on the basis of protein target mutations, the target remains vulnerable and the organism susceptible if access to the target can be restored. For this reason, chemosensitizers (or so-called 'resistance-reversal agents') that interact with PfCRT to 'reverse' quinoline resistance have been studied, but have largely failed to gain traction as candidate components of antimalarial combinations. Beyond the challenge of achieving adequate potency and safety, quinoline chemosensitizers have also lacked intrinsic antimalarial efficacy, and thus when combined with a quinoline antimalarial, would effectively result in monotherapy[14–16]. Because of the critical value of haem detoxification as a drug target and the appeal of preserving or restoring quinoline efficacy, we have sought to develop an intrinsically active antimalarial that also restores quinoline sensitivity to multi-drug resistant (MDR) parasites.

Exploiting the acridone chemical structure, we engineered a new scaffold with both features incorporated into one molecule: (1) a haem-targeting tricyclic mainframe with an ionizable side chain to promote accumulation in the digestive vacuole, and (2) a chemosensitization moiety at the N10 position to counteract quinoline resistance (Fig. 1). The rigid tricyclic aromatic acridone core promotes π–π stacking for haem binding. The side chain at position 6 engages one of the propionates of haem in an ionic bond and promotes acid trapping in the digestive vacuole. The side-chain attachment at the central nitrogen atom provides a hydrogen bond acceptor needed for the chemosensitization function. This feature is a well-established component of the pharmacophore for effective chloroquine chemosensitizers[17,18], including previously described acridone chemosensitizers[16], and further confirmed by the ineffectiveness as a quinoline chemosensitizer of T2 (3-chloro-6-(2-diethylamino-ethoxy)-10*H*-acridone), an intrinsically potent acridone derivative lacking the essential side chain attached to the central nitrogen atom (Supplementary Table 1).

Twelve compounds synthesized using this rational design approach were initially screened for intrinsic activity (Supplementary Table 2). *In vitro* antimalarial potency was demonstrated against a panel of chloroquine-sensitive and MDR strains of *P. falciparum* with different geographic and genetic backgrounds. Results from testing of our initial lead compound T3.5 are shown in Table 1. *In vivo* intrinsic antimalarial efficacy of T3.5 against patent infection in mice was demonstrated using once-daily oral dosing for 3 days in two murine models. A dose

[1]Portland Veterans Affairs Medical Centre, Portland, Oregon 97239, USA. [2]Department of Chemistry, Portland State University, Portland, Oregon 97201, USA. [3]Oregon Health and Science University, Portland, Oregon 97239, USA. [4]Swiss Tropical Institute, Socinstrasse 57, CH-4002 Basel, Switzerland. [5]Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA. †Present addresses: Oregon Translational Research and Drug Development Institute, Portland, Oregon 97201, USA (M.J.S.); Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, Virginia 23298, USA (K.D.L.).

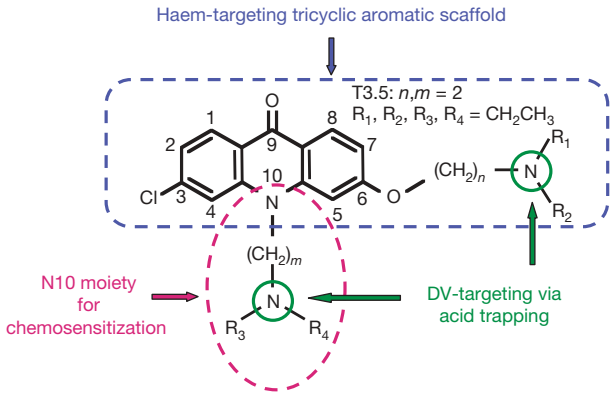


Figure 1 | Generalized chemical structure of dual-function acridone derivatives. The rigid tricyclic aromatic acridone core promotes π–π stacking for haem binding. The side-chain attachment at the central nitrogen atom provides a hydrogen bond acceptor needed for the chemosensitization function, and, together with the side chain at position 6, facilitates accumulation in the digestive vacuole (DV) via acid trapping.

of 100 mg kg^{-1} day^{-1} T3.5 diminished *Plasmodium berghei* parasitemia by 95%; against *Plasmodium yoelii*, T3.5 dose-response testing showed half-maximal effective dose (ED_{50}) and ED_{90} values of 56 and 88 mg kg^{-1} day^{-1}, respectively (Table 2). Initial high-dose testing (256 mg kg^{-1} day^{-1} orally or 200 mg kg^{-1} day^{-1} intraperitoneally) was curative. No overt toxicity or behaviour change was observed in the assessment of general measures of animal well-being (weight, grooming or locomotor activity), and there has been no apparent *in vitro* mammalian cell cytotoxicity against the proliferation of murine splenic lymphocytes or human foreskin fibroblast cells (Supplementary Table 3). Apparent structural similarities between T3.5 and tricyclic antidepressants led us to assess the activity of T3.5 in a model of cloned biogenic amine transporters. Unlike cyclic antidepressants, T3.5 had no significant affinity with serotonin, dopamine or noradrenaline transporters (Supplementary Table 4).

The *in vitro* interactions of T3.5 with other antimalarials were assessed using a rigorous fixed-ratio combination strategy[16,19]. In combination with five prototypical quinoline derivatives, T3.5 proved synergistic with chloroquine, amodiaquine, quinine or piperaquine, but not with mefloquine, against the MDR *P. falciparum* strain Dd2 (Fig. 2a, b). As illustrated by the isobologram in Fig. 2a, there was no synergy in the additive interaction between T3.5 and chloroquine against the chloroquine-sensitive parasite D6. In contrast, the synergy with quinine is distinct. Classic verapamil-like quinoline chemosensitizers (including earlier acridones[16]) modulate drug sensitivity only in drug-resistant parasites, the effect is more pronounced in the 'Old World' (Asian/African) phenotype than in the 'New World' (American/Oceanic) phenotype, and often micromolar concentrations are required[14–16]. As shown in Fig. 2c (solid line), the T3.5 and quinine combination is entirely different, demonstrating equal synergy in both Dd2 (Indochina) and 7G8 (Brazil) strains of *P. falciparum*, and more remarkably, synergy against quinine-sensitive D6 (Africa) (mean fractional inhibitory concentration (FIC) indices 0.50, 0.49 and 0.64, respectively). Notably, similar synergy characteristics

Table 1 | Intrinsic *in vitro* antimalarial activity against *P. falciparum*

Compound	IC_{50} (nM) versus *P. falciparum*			
	D6*	Dd2*	7G8*	Tm90-C2B*
T3.5	44.8 ± 5.2	77.3 ± 6.0	85.9 ± 6.8	71.3 ± 7.4
Chloroquine	8.4 ± 1.9	124.7 ± 9.9	235.7 ± 25.6	122.7 ± 10.5
Quinine	19.4 ± 1.1	87.2 ± 9.7	77.3 ± 6.9	55.8 ± 5.5

Values are the mean ± s.e.m. from eight independent experiments, each in quadruplicate, using an MSF assay with 0.2% parasitemia and 2% hematocrit.
* D6 (Africa): chloroquine-sensitive; Dd2 (Indochina): MDR; 7G8 (Brazil): MDR; Tm90-C2B (Thailand): MDR (including atovaquone and anti-folate).

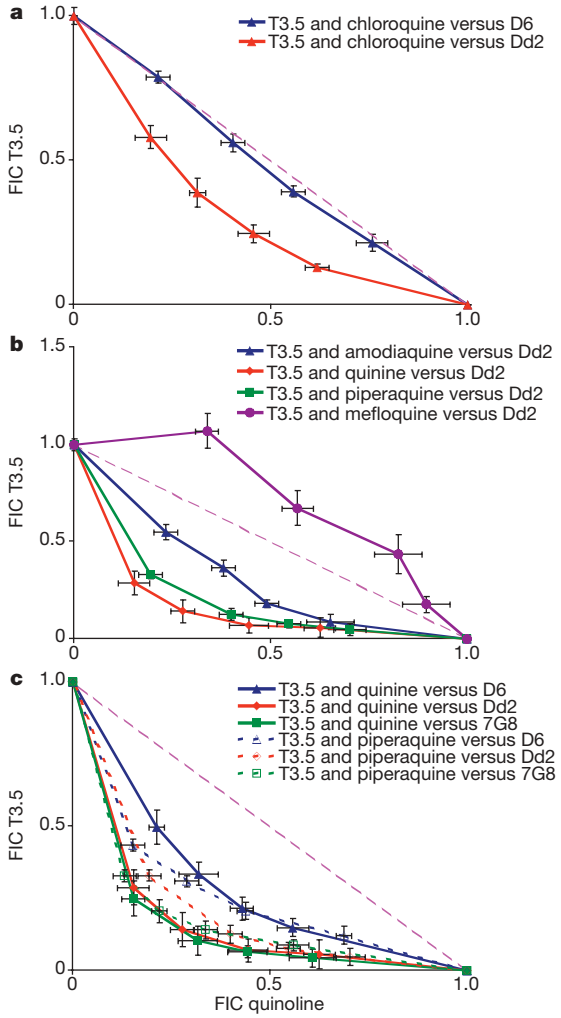


Figure 2 | The *in vitro* interactions of T3.5 with other antimalarials. **a**–**c**, Isobolograms of the *in vitro* interaction of (**a**) the T3.5 and chloroquine combination against MDR *P. falciparum* strain Dd2, and against chloroquine-sensitive *P. falciparum* strain D6 (mean FIC indices are 0.72 and 0.97, respectively); (**b**) the T3.5 and amodiaquine, T3.5 and mefloquine, T3.5 and quinine, and T3.5 and piperaquine combinations against MDR strain Dd2 (mean FIC indices are 0.73, 1.25, 0.50 and 0.60, respectively); (**c**) the T3.5 and quinine combination (solid lines) against chloroquine-sensitive strain D6, and against MDR strains Dd2 and 7G8 (mean FIC indices are 0.64, 0.50 and 0.49, respectively); and the T3.5 and piperaquine combination (dashed lines) against chloroquine-sensitive strain D6, and against MDR strains Dd2 and 7G8 (mean FIC indices are 0.66, 0.60 and 0.51, respectively). The mean FIC indices ± s.e.m. were derived from three independent experiments. The *x* axes represents the FICs of quinoline, and the *y* axes represents the FICs of T3.5. The diagonal line (FIC index = 1) indicates the hypothetical additive drug effect. A concave curve (FIC index < 1) below the diagonal line typically indicates synergy of the combination, whereas a convex curve (FIC index > 1) above the diagonal line indicates antagonism.

were also observed for the T3.5 and piperaquine combination (Fig. 2c, dashed line).

The synergy between T3.5 and quinine was also observed *in vivo* against patent infection with quinine-sensitive *P. yoelii*. In combination, substantial reductions in the effective dosage of T3.5 and quinine were noted. For example, the ED_{90} values for either T3.5 or quinine alone were 88 and 85 mg kg^{-1} day^{-1}, respectively, but the same effect was achieved by combining less than 1/3 of those individual doses (Table 2).

The profile of the interaction between T3.5 and PfCRT was assessed using PfCRT mutant lines with various point mutations at

Table 2 | Synergism of T3.5 and quinine combination against *P. yoelii*

Effect	Dose (mg kg^{-1} day^{-1})		
	T3.5 alone	Quinine alone	T3.5 and quinine combination
ED_{50}	56 ± 7	39 ± 4	14:14 ± 4:4
ED_{75}	70 ± 11	57 ± 9	19:19 ± 4:4
ED_{90}	88 ± 20	85 ± 25	24:24 ± 7:7

Values are the mean ± s.d.

codon 76, previously used to distinguish characteristics of verapamil-like chemosensitizers[12,13,20–22]. With chloroquine, the interaction with T3.5 mirrored that of verapamil; however, the chemosensitization pattern of T3.5 with quinine clearly differs (Table 3). In contrast to verapamil-like chemosensitizers that are more intrinsically potent against *P. falciparum* with a K76N mutation, the intrinsic efficacy of T3.5 is comparable against all four PfCRT mutant lines. Unlike verapamil, T3.5 is synergistic with quinine against the quinine-hypersensitive mutant line 106/1^{76I}. Moreover, in contrast to verapamil, T3.5 potentiated quinine activity against the chloroquine-sensitive parent line Sudan 106/1^{K76}. In fact, T3.5 demonstrated chemosensitization to quinine against more than two dozen strains of *P. falciparum* with wide variations in quinoline-resistance profiles (data not shown). Such broad synergistic interaction with quinine has not, to our knowledge, been previously described.

Investigations of the proposed mechanisms of intrinsic activity confirmed the interaction of T3.5 with haem, interference with haemozoin formation, and drug accumulation within the digestive vacuole. Examination of Giemsa-stained blood smears after T3.5 exposure revealed a dose-related failure of parasite progression to schizogeny, and markedly diminished haemozoin content in *P. yoelii*-infected mice and in T3.5-exposed *P. falciparum* (D6 and Dd2) *in vitro* (Supplementary Fig. 1). Quantitative measurement of haemozoin production correlates well with microscopy (Supplementary Fig. 2). Intracellular haemozoin incorporated during 24 h of incubation with 500 nM T3.5 decreased from 0.78 to 0.06 fmol per parasitized erythrocyte in the D6 strain, and from 1.09 to 0.25 fmol in the Dd2 strain. In addition, T3.5 inhibits β-hematin formation *in vitro* (Supplementary Table 5) and interacts with haem to form a soluble complex with strong affinity at pH 5.2 (Supplementary Fig. 3). Visual localization of T3.5 inside the *P. falciparum*-infected erythrocyte by confocal fluorescence microscopy indicates uptake and accumulation in the digestive vacuole (Fig. 3 and Supplementary Methods).

There is current consensus that combination therapy is essential to delay the emergence of drug resistance and preserve the long-term usefulness of new antimalarials[23]. The search for alternatives to artemisinin-based combinations has increased interest in developing 'dual-function' antimalarials with both intrinsic and chemosensitizing efficacy[24,25]. Some existing digestive vacuole-active compounds have intrinsic efficacy against *Plasmodium* parasites, regardless of the chloroquine-resistance profile. Because these compounds probably interact or interfere with PfCRT, there has been speculation that they might also exhibit chemosensitization in combination with quinolines[26,27], but no such activity has been demonstrated. The present study provides, to our knowledge, the first evidence of such dual functionality in a single molecule, and offers a powerful approach to combination therapy.

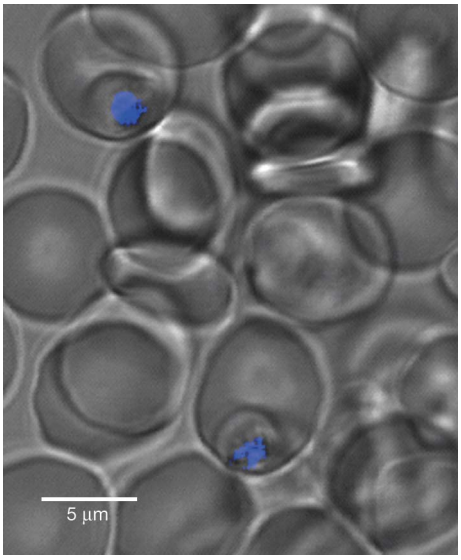


Figure 3 | Confocal microscopy of localized T3.5 fluorescence in two intraerythrocytic *P. falciparum* trophozoites. Fluorescence was determined in live *P. falciparum*-infected erythrocytes using a laser (emission line 351 nm) scanning confocal microscope. The figure shows the intrinsic fluorescence of T3.5 (blue) superimposed on the bright-field transmission image of the infected cells. Scale bar, 5 μm.

If a global effort to eradicate malaria is to be successful, the drug therapy component of that effort must address the gaps and weaknesses in the armamentarium of the therapies available at present. Affordability, safety in the most vulnerable, and low susceptibility to drug-resistance adaptations each represent unmet needs. In contrast to other drug classes (such as respiratory inhibitors and anti-folates), the development of drug resistance to digestive vacuole-active drugs that target haem-processing has been slow (chloroquine) or of low order (quinine), and this remains the only identified immutable parasite target. For older drugs in this class, cost is very low, there is extensive experience with their use in children and during pregnancy, and short-course therapy is facilitated by very long elimination times. Paradoxically then, although the failure of chloroquine is at the core of the global drug-resistance crisis, these drugs actually characterize the ideals now sought in new antimalarial drugs for both treatment and intermittent prophylaxis. The concept described in this paper specifically aims to exploit the strengths of such compounds by making possible a new combination therapy strategy. The ability to maintain the efficacy of newer drugs (such as piperaquine) and to restore the efficacy of older drugs (such as chloroquine) represents a uniquely powerful tool, and one ideally suited to achieve the broadest possible benefit as a renewed malaria eradication effort proceeds.

Table 3 | Effect of PfCRT position 76 mutations on chemosensitization in *P. falciparum*

Compound	IC_{50} (nM) *P. falciparum*			
	106/1^{K76}*	106/1^{76T}†	106/1^{76N}†	106/1^{76I}†
Verapamil alone	25,800 ± 128	1,779 ± 105	632 ± 78	8630 ± 312
T3.5 alone	93.6 ± 12	86.9 ± 8.5	134 ± 15	100.8 ± 18
Chloroquine alone	15.9 ± 1.4	188 ± 2.3	147 ± 6.8	158 ± 4.4
Chloroquine + 500 nM verapamil	17.2 ± 1.6	51.2 ± 3.2	39.9 ± 5.4	47.7 ± 5.6
Chloroquine + 25 nM T3.5	14.5 ± 2.1	67.8 ± 9.8	74.3 ± 6.8	58.5 ± 6.1
Quinine alone	112 ± 18	204 ± 19	186 ± 21	20.8 ± 2.1
Quinine + 500 nM verapamil	118 ± 10	71.4 ± 8.9	53.5 ± 4.6	32.1 ± 2.4
Quinine + 25 nM T3.5	55.6 ± 4.0	56.7 ± 5.3	71.2 ± 6.8	5.5 ± 0.8

Values are the mean ± s.e.m. from three independent experiments, each in quadruplicate, using the MSF assay with 0.2% parasitemia and 2% hematocrit.
* Sudan 106/1^{K76} is known to have six of the seven mutations in the *pfcrt* gene necessary for chloroquine resistance.
† Mutant line selected with chloroquine pressure from 106/1^{K76}.

METHODS SUMMARY

In vitro antimalarial activity was determined by the Malaria SYBR Green I-based Fluorescence (MSF) assay described previously[28] with slight modification[16]. *In vivo* efficacy was determined using once-daily oral dosing for 3 days against patent infection in two murine models. *In vitro* interaction of T3.5 and other antimalarial agents was assessed by isobolar analysis using fixed-ratio combination, in which drugs were diluted in fixed ratios of starting concentrations predetermined to generate well-defined concentration response curves[16,19]. The effect of PfCRT mutations on drug interaction was determined by a modified MSF method described previously[13,16].

Received 18 December 2008; accepted 25 February 2009.
Published online 8 April 2009.

1. Yayon, A., Cabantchik, Z. I. & Ginsburg, H. Identification of the acidic compartment of *Plasmodium falciparum*-infected human erythrocytes as the target of the antimalarial drug chloroquine. *EMBO J.* **3,** 2695–2700 (1984).
2. Olliaro, P. L. & Goldberg, D. E. The *Plasmodium* digestive vacuole: metabolic headquarters and choice drug target. *Parasitol. Today* **11,** 294–297 (1995).
3. Sullivan, D. J. Jr, Gluzman, I. Y., Russell, D. G. & Goldberg, D. E. On the molecular mechanism of chloroquine's antimalarial action. *Proc. Natl Acad. Sci. USA* **93,** 11865–11870 (1996).
4. Wellems, T. E. *Plasmodium* chloroquine resistance and the search for a replacement antimalarial drug. *Science* **298,** 124–126 (2002).
5. Kumar, S., Guha, M., Choubey, V., Maity, P. & Bandyopadhyay, U. Antimalarial drugs inhibiting hemozoin (β-hematin) formation: a mechanistic update. *Life Sci.* **80,** 813–828 (2007).
6. Krugliak, M., Zhang, J. & Ginsburg, H. Intraerythrocytic *Plasmodium falciparum* utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol for the biosynthesis of its proteins. *Mol. Biochem. Parasitol.* **119,** 249–256 (2002).
7. Pagola, S., Stephens, P. W., Bohle, D. S., Kosar, A. D. & Madsen, S. K. The structure of malaria pigment β-haematin. *Nature* **404,** 307–310 (2000).
8. Egan, T. J. Haemozoin formation. *Mol. Biochem. Parasitol.* **157,** 127–136 (2008).
9. Foley, M. & Tilley, L. Quinoline antimalarials: mechanisms of action and resistance. *Int. J. Parasitol.* **27,** 231–240 (1997).
10. Fitch, C. D. Chloroquine resistance in malaria: a deficiency of chloroquine binding. *Proc. Natl Acad. Sci. USA* **64,** 1181–1187 (1969).
11. Krogstad, D. J. *et al.* Efflux of chloroquine from *Plasmodium falciparum*: mechanism of chloroquine resistance. *Science* **238,** 1283–1285 (1987).
12. Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6,** 861–871 (2000).
13. Cooper, R. A. *et al.* Alternative mutations at position 76 of the vacuolar transmembrane protein PfCRT are associated with chloroquine resistance and unique stereospecific quinine and quinidine responses in *Plasmodium falciparum*. *Mol. Pharmacol.* **61,** 35–42 (2002).
14. Martin, S. K., Oduola, A. M. & Milhous, W. K. Reversal of chloroquine resistance in *Plasmodium falciparum* by verapamil. *Science* **235,** 899–901 (1987).
15. van Schalkwyk, D. A. & Egan, T. J. Quinoline-resistance reversing agents for the malaria parasite *Plasmodium falciparum*. *Drug Resist. Updat.* **9,** 211–226 (2006).
16. Kelly, J. X. *et al.* Design, synthesis, and evaluation of 10-N-substituted acridones as novel chemosensitizers in *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **51,** 4133–4140 (2007).
17. Bhattacharjee, A. K., Kyle, D. E. & Vennerstrom, J. L. Structural analysis of chloroquine resistance reversal by imipramine analogs. *Antimicrob. Agents Chemother.* **45,** 2655–2657 (2001).
18. Bhattacharjee, A. K., Kyle, D. E., Vennerstrom, J. L. & Milhous, W. K. A 3D QSAR pharmacophore model and quantum chemical structure—activity analysis of chloroquine(CQ)-resistance reversal. *J. Chem. Inf. Comput. Sci.* **42,** 1212–1220 (2002).
19. Fivelman, Q. L., Adagu, I. S. & Warhurst, D. C. Modified fixed-ratio isobologram method for studying in vitro interactions between atovaquone and proguanil or dihydroartemisinin against drug-resistant strains of *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **48,** 4097–4102 (2004).
20. Waller, K. L. *et al.* Chloroquine resistance modulated in vitro by expression levels of the *Plasmodium falciparum* chloroquine resistance transporter. *J. Biol. Chem.* **278,** 33593–33601 (2003).
21. Cooper, R. A., Hartwig, C. L. & Ferdig, M. T. *pfcrt* is more than the *Plasmodium falciparum* chloroquine resistance gene: a functional and evolutionary perspective. *Acta Trop.* **94,** 170–180 (2005).
22. Cooper, R. A. *et al.* Mutations in transmembrane domains 1, 4 and 9 of the *Plasmodium falciparum* chloroquine resistance transporter alter susceptibility to chloroquine, quinine and quinidine. *Mol. Microbiol.* **63,** 270–282 (2007).
23. World Health Organization. *Guidelines for the Treatment of Malaria. WHO/HTM/MAL/2006.1108* (WHO, 2006).
24. Biot, C. & Chibale, K. Novel approaches to antimalarial drug discovery. *Infect. Disord. Drug Targets* **6,** 173–204 (2006).
25. Egan, T. J. & Kaschula, C. H. Strategies to reverse drug resistance in malaria. *Curr. Opin. Infect. Dis.* **20,** 598–604 (2007).
26. Biot, C. *et al.* Insights into the mechanism of action of ferroquine. Relationship between physicochemical properties and antiplasmodial activity. *Mol. Pharm.* **2,** 185–193 (2005).
27. Burgess, S. J. *et al.* A chloroquine-like molecule designed to reverse resistance in *Plasmodium falciparum*. *J. Med. Chem.* **49,** 5623–5625 (2006).
28. Smilkstein, M., Sriwilaijaroen, N., Kelly, J. X., Wilairat, P. & Riscoe, M. Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. *Antimicrob. Agents Chemother.* **48,** 1803–1806 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Kyle for the gift of *P. falciparum* parasite Tm90-C2B, and the Malaria Research and Reference Resource Center for supplying *P. falciparum* parasites D6, Dd2 and 7G8. We thank A. P. Waters and C. J. Janse for the donation of GFP-labelled *P. berghei* strain ANKA. We thank L. Jones-Brando and R. Yolken for the cytotoxicity (HFF) data. We are grateful to A. Cornea for the confocal imaging. We thank K. Liebman, C. Hudson, S. Burgess and D. Peyton for compound characterization. We acknowledge financial support from the Merit Review Program of the Department of Veterans Affairs. United States patent applications have been filed by the US Department of Veterans Affairs to protect the intellectual property described in this report.

Author Contributions J.X.K. contributed to drug design, synthesis, *in vitro* screening, *in vitro* drug combination studies, haemozoin inhibition assays, and manuscript preparation. M.J.S. contributed to *in vivo* testing in the *P. yoelii* murine model, *in vitro* drug combination studies, confocal imaging, and manuscript preparation. R.B. and S.W. contributed to *in vivo* testing in the *P. berghei* murine model. R.A.C. and K.D.L. contributed to examination of the PfCRT mutant lines. A.J. and R.A.J. contributed to biogenic amine studies. R.W. and R.A.D. assisted with synthesis. D.J.H. assisted with *in vivo* screening in the *P. yoelii* murine model. M.K.R. contributed to drug design and review of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.X.K. (kellyja@ohsu.edu) or M.K.R. (riscoem@ohsu.edu).

LETTERS

qiRNA is a new type of small interfering RNA induced by DNA damage

Heng-Chi Lee[1], Shwu-Shin Chang[1], Swati Choudhary[1], Antti P. Aalto[2], Mekhala Maiti[1]†, Dennis H. Bamford[2] & Yi Liu[1]

RNA interference pathways use small RNAs to mediate gene silencing in eukaryotes. In addition to small interfering RNAs (siRNAs) and microRNAs, several types of endogenously produced small RNAs have important roles in gene regulation, germ cell maintenance and transposon silencing[1–4]. The production of some of these RNAs requires the synthesis of aberrant RNAs (aRNAs) or pre-siRNAs, which are specifically recognized by RNA-dependent RNA polymerases to make double-stranded RNA. The mechanism for aRNA synthesis and recognition is largely unknown. Here we show that DNA damage induces the expression of the Argonaute protein QDE-2 and a new class of small RNAs in the filamentous fungus *Neurospora crassa*. This class of small RNAs, known as qiRNAs because of their interaction with QDE-2, are about 20–21 nucleotides long (several nucleotides shorter than *Neurospora* siRNAs), with a strong preference for uridine at the 5′ end, and originate mostly from the ribosomal DNA locus. The production of qiRNAs requires the RNA-dependent RNA polymerase QDE-1, the Werner and Bloom RecQ DNA helicase homologue QDE-3 and dicers. qiRNA biogenesis also requires DNA-damage-induced aRNAs as precursors, a process that is dependent on both QDE-1 and QDE-3. Notably, our results suggest that QDE-1 is the DNA-dependent RNA polymerase that produces aRNAs. Furthermore, the *Neurospora* RNA interference mutants show increased sensitivity to DNA damage, suggesting a role for qiRNAs in the DNA-damage response by inhibiting protein translation.

In the filamentous fungus *Neurospora crassa*, the RNA interference (RNAi) pathway is essential for both double-stranded RNA (dsRNA)-induced and transgene-induced gene silencing (quelling)[5]. In the quelling pathway, QDE-1 and QDE-3 are thought to be involved in the generation of dsRNA[6,7]. Furthermore, QDE-3 was previously shown to be involved in DNA repair[6]. It has been proposed that a repetitive transgene leads to the production of transgene-specific aRNA, which is converted to dsRNA by QDE-1. Two partially redundant Dicer proteins, DCL-1 and DCL-2, cleave the dsRNA into siRNAs of around 25 nucleotides in size[8]. Subsequently, the siRNAs are loaded onto the RNA-induced silencing complex (RISC), formed by the Argonaute protein QDE-2 and an exonuclease QIP[9–11]. We previously showed that dsRNA, but not siRNA, transcriptionally activates *qde-2*, other RNAi components and putative antiviral genes[12].

During our study of QDE-2 regulation, we observed that supplementing histidine in the medium resulted in a significant increase of *qde-2* messenger RNA and QDE-2 protein levels, whereas the addition of other amino acids did not (Fig. 1a and Supplementary Fig. 1a). Histidine is known to inhibit DNA replication, reduce the nucleoside 5′-triphosphate pool, and result in DNA damage in *Neurospora*[13,14]. In addition, histidine significantly increased the mutation rate at the *mtr* locus (Supplementary Fig. 1b). These results suggest that DNA damage results in the induction of *qde-2* expression. Treatment with ethyl methanesulphonate (EMS, Fig. 1b), hydroxyurea (Supplementary Fig. 1c), or methyl methanesulphonate (data not shown) also induced QDE-2 expression. Notably, the induction of QDE-2 by histidine and other DNA-damaging agents requires QDE-1, QDE-3 and the DCLs (Fig. 1b and Supplementary Fig. 1c, d). Furthermore, in the absence of DNA damaging agents, QDE-2 accumulates to increased levels in DNA repair mutants that are deficient in dsDNA break repair or homologous recombination repair pathways (Fig. 1c). Taken together, these results demonstrate that DNA damage activates *qde-2* expression.

Because QDE-1 and QDE-3 are involved in the generation of dsRNA, and DCLs are important for maintaining the steady state levels of QDE-2 post-transcriptionally[12], our results indicate that DNA damage results in the production of endogenous dsRNA, which activates *qde-2* transcription. We reasoned that such dsRNA is processed into small RNAs, which then associate with QDE-2. To examine this possibility, we immunoprecipitated c-Myc-tagged QDE-2 expressed in a *qde-2* null strain[11]. The QDE-2-associated RNA was extracted and 3′-end labelled with ^{32}P-cytidine bisphosphate. As shown in Fig. 2a, Myc–QDE-2 specifically associated with a group of small RNAs

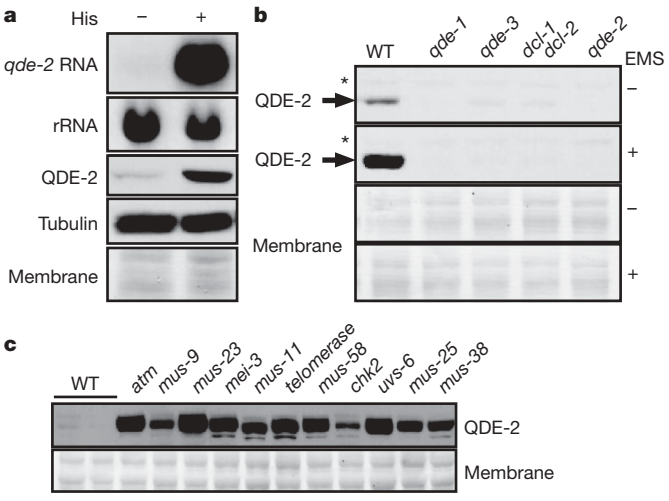


Figure 1 | DNA damage induces QDE-2 expression. a, Northern and western blot analyses showing the induction of *qde-2* mRNA and QDE-2 protein expression by histidine. **b**, Western blot analysis showing that the induction of QDE-2 by EMS requires QDE-1, QDE-3 and DCLs. The labels above the lanes indicate the genotypes of the strains. Asterisks indicate a nonspecific protein band recognized by the QDE-2 antibody. WT, wild type. **c**, Western blot analyses showing high QDE-2 levels in DNA repair mutants in the absence of a DNA damaging agent.

[1]Department of Physiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA. [2]Institute of Biotechnology and Department of Biological and Environmental Sciences, Biocenter 2, PO Box 56, FIN-00014 University of Helsinki, Helsinki, Finland. †Present address: Department of Cell Biology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA.

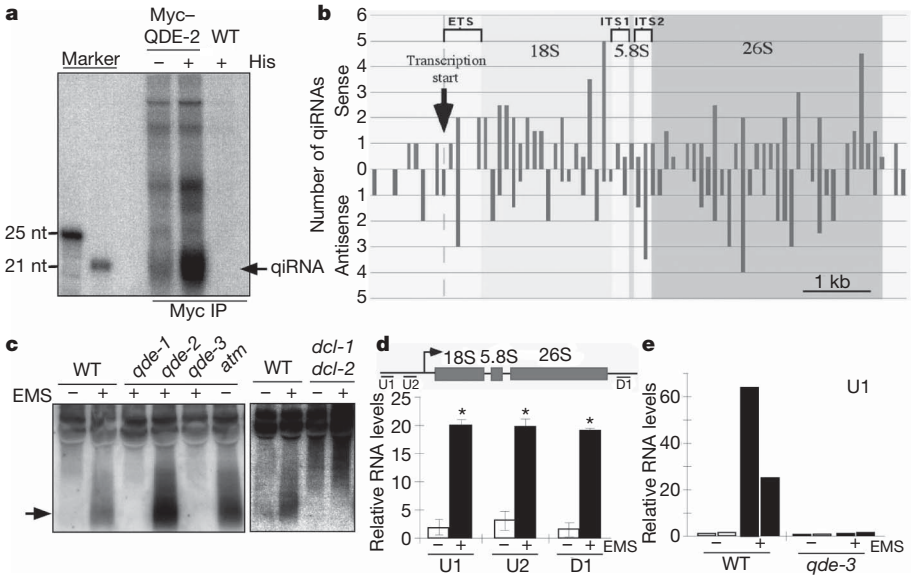


Figure 2 | DNA damage results in the production of qiRNAs and aRNAs. **a**, Enrichment of the Myc–QDE-2-associated small RNAs by immunoprecipitation (IP). A wild-type (WT) strain that lacks the Myc–QDE-2 construct was used as a negative control. Immunoprecipitated RNA was 3′-end labelled and separated on a 16% acrylamide gel. nt, nucleotides. **b**, Mapping of qiRNAs to an rDNA repeat. Regions encoding the mature rRNAs are shaded. **c**, Northern blot analysis of qiRNAs in the indicated strains. An RNA probe specific for the antisense 26S RNA was used. The arrow denotes the qiRNAs. **d**, qRT–PCR showing that EMS treatment results in the induction of aRNAs from the rDNA regions. The top panel shows a schematic diagram of an rDNA repeat and the intergenic rDNA regions (U1, U2 and D1) analysed by qRT–PCR analysis. A *dcl* double mutant was used. $n = 3$; $^*P < 0.001$; error bars indicate s.d. **e**, qRT–PCR analysis showing the loss of EMS-induced aRNAs from the rDNA locus in the *qde-3* mutant. Results of two independent experiments are shown.

approximately 20–21 nucleotides in length, which were markedly induced after histidine or EMS treatment (data not shown). Because these small RNAs are endogenously produced and are associated with QDE-2, they were named qiRNAs for QDE-2-interacting small RNAs. The average length of qiRNAs is several nucleotides shorter than *Neurospora* siRNAs (Fig. 2a and Supplementary Fig. 2a), which are around 25 nucleotides[8].

qiRNAs were cloned and sequenced. Analyses of 184 individual qiRNA sequences showed that they indeed possess an average length of about 20–21 nucleotides (Supplementary Fig. 2b and Supplementary Table 1). Similar to the Piwi-interacting RNAs (piRNAs) recently identified in animals[15], the first nucleotide of the 5′ end of qiRNAs exhibits a strong preference for uracil (93%) (Supplementary Fig. 2c). Also, the first nucleotide of the 3′ end prefers adenine (49%).

Surprisingly, most qiRNAs (86%) originated from the ribosomal DNA (rDNA) locus (Supplementary Fig. 2d), where ~200 copies of rDNA repeats form the nucleolus organizer region. Their association with QDE-2 and their 5′ and 3′ end nucleotide preferences suggest that qiRNAs are not nonspecific ribosomal RNA degradation products. The remaining qiRNAs were mapped to intergenic regions (6.57%), open reading frames (4.37%) and transfer RNAs (1.45%).

qiRNAs from the rDNA locus correspond to both sense and antisense strands at approximately equal frequency, suggesting that the biogenesis of qiRNAs requires the formation of dsRNA (Fig. 2b). Furthermore, qiRNAs not only originate from the region corresponding to the mature rRNAs, but many derive from the external and internal transcribed spacer regions (ETS, ITS1 and ITS2) and the intergenic spacer regions. These results indicate that the biogenesis of qiRNAs may require unconventional transcriptional events.

Northern blot analysis showed that the levels of 26S rDNA-specific qiRNA were undetectable under normal conditions but were markedly induced after EMS treatment (Fig. 2c). qiRNA also accumulated to a high level in an *atm* mutant without EMS treatment. Furthermore, qiRNAs production was completely abolished in the *qde-1* and *qde-3* mutant strains (Fig. 2c), indicating that QDE-1 and QDE-3 are required for qiRNA biogenesis. In contrast, the production of qiRNA was maintained in the *qde-2* mutant. Although the sizes of qiRNAs are smaller than those of siRNAs, the production of qiRNA is abolished in a *dcl-1 dcl-2* double mutant. Moreover, long RNA species accumulated in the *dcl* double mutant after DNA damage, suggesting the accumulation of long dsRNA.

To investigate the relationship between qiRNAs and aRNAs, we examined the transcript levels from the intergenic rDNA spacer regions. Quantitative PCR with reverse transcription (qRT–PCR) and northern blot analyses showed that the RNA transcripts originating from both upstream and downstream regions of the transcribed rDNA region are indeed highly induced after DNA damage (Fig. 2d and Supplementary Fig. 3). In the *dcl* double mutant, aRNAs accumulated to a high level, with sizes ranging from a few hundred nucleotides to ~2 kilobases (kb) (Supplementary Fig. 3), suggesting that these transcripts form dsRNA. Notably, we found that aRNA production was completely abolished in the *qde-3* mutant (Fig. 2e), indicating that aRNAs are the precursors of qiRNAs and that QDE-3, the RecQ DNA helicase, is required for aRNA biogenesis.

RNA polymerase I is responsible for the transcription of rRNAs. However, we found that the DNA-damage-induced aRNA production was maintained in an RNA polymerase I mutant (Supplementary Fig. 4a–c). Furthermore, both qRT–PCR and northern blot analyses showed that the treatment of *Neurospora* with thiolutin, a potent inhibitor of RNA polymerases I, II and III, did not affect aRNA production despite its strong inhibition of mRNA synthesis (Fig. 3a and Supplementary Fig. 5a). These results suggest that the common RNA polymerases are not required for the generation of aRNAs after DNA damage.

QDE-1, the RNA-dependent RNA polymerase (RdRP), was previously thought to specifically recognize and convert aRNA into dsRNA. To our surprise, we found that the induction of rDNA-specific aRNAs by histidine was completely abolished in the *qde-1*^{ko} mutant (Fig. 3b and Supplementary Fig. 5b), indicating that QDE-1 is not only required for the production of dsRNA but also for the synthesis of aRNAs. Moreover, we found that QDE-1 does not amplify small RNAs in dsRNA-mediated gene silencing (Supplementary Fig. 5c, d), suggesting that qiRNAs are not amplification products of primary small RNAs.

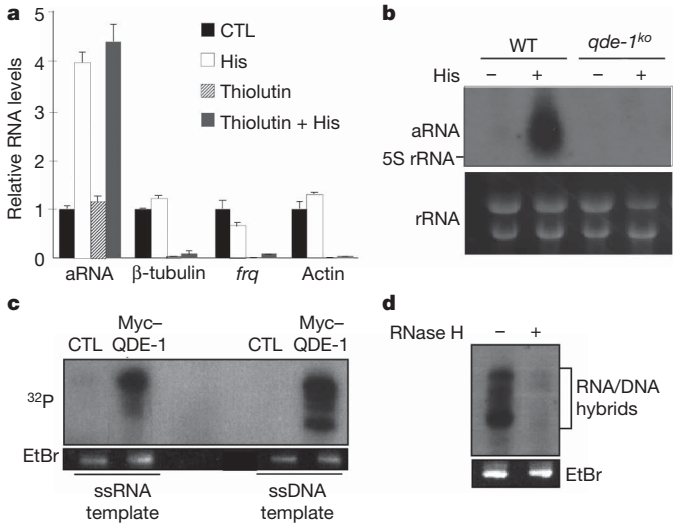


Figure 3 | QDE-1 is required for the synthesis of DNA-damage-induced rDNA-specific aRNA and exhibits DdRP activity using an ssDNA template. **a**, qRT–PCR results showing the levels of the rDNA-specific aRNA and several Pol-II-transcribed genes in the wild-type strain. CTL denotes the control sample. $n = 3$; error bars indicate s.d. **b**, Northern blot analysis showing the levels of aRNA in the wild-type (WT) and *qde-1*^{ko} strains. Total RNA was used. **c**, *In vitro* RNA polymerase assay using a 175-nucleotide ssDNA or ssRNA template and purified Myc–QDE-1. CTL indicates a reaction using purification products from a strain without the construct. EtBr, ethidium bromide. **d**, The ssDNA-templated reaction products from **c** were treated with RNase H.

Recent structural analysis of QDE-1 has shown that its catalytic core is similar to eukaryotic DNA-dependent RNA polymerases (DdRPs) but not to viral RdRPs[16]. This result prompted us to examine whether QDE-1 can function as a DdRP to generate aRNAs. A c-Myc-His-tagged QDE-1 was expressed in the *qde-1*^{ko} strain and purified by affinity purification to be used in RdRP and DdRP assays. As shown in Fig. 3c, QDE-1 exhibited RNA polymerase activity using both single-stranded RNA (ssRNA) and ssDNA as templates to generate full-length RNA products. In contrast, RNA polymerase activity was not detected using the control purification products. In addition, RNase H degraded the ^{32}P-labelled ssDNA-templated products of QDE-1 (Fig. 3d), indicating that they were mostly DNA/RNA hybrids. We also found that the RNA polymerase activity of QDE-1 is not inhibited by thiolutin *in vitro* (H.-C.L., A.P.A., D.H.B. and Y.L., unpublished observations). Together, these results demonstrate that QDE-1 can function as both an RdRP and a DdRP. The requirement of QDE-1 for aRNA synthesis suggests that QDE-1 is the RNA polymerase that generates aRNA.

To determine whether the rDNA-specific qiRNAs are functional, we immunoprecipitated Myc–QDE-2 using strains that express either wild-type QDE-2 or QDE-2 containing a D664A mutation that abolishes its catalytic activity[11]. As shown in Fig. 4a, qiRNAs were associated with both forms of QDE-2. However, the qiRNAs associated with wild-type QDE-2 were entirely single-stranded, whereas only double-stranded qiRNAs bound to QDE-2(D664A). This result indicates that qiRNAs are associated with an active RISC complex.

Because most qiRNAs are derived from the rDNA locus, they may inhibit rRNA biogenesis and protein synthesis after DNA damage. As shown in Fig. 4b, the protein synthesis rate measured by a ^{35}S-labelling pulse (^{35}S-Met and ^{35}S-Cys) was significantly decreased after histidine treatment. Notably, this decrease in protein synthesis rate was partially blocked in the *qde-1* and *qde-3* mutants ($P = 4.6 \times 10^{-6}$ and 2.2×10^{-5}, respectively). Similar results were also obtained using EMS (Supplementary Fig. 6a). These results suggest that qiRNAs are involved in inhibiting protein synthesis after DNA damage.

Consistent with a role for qiRNAs in the DNA-damage response, a *qde-3* mutant was previously shown to be sensitive to both histidine and DNA damaging agents[6]. Furthermore, we found that both *qde-1* and the *dcl* double mutants had increased sensitivity to histidine, EMS and hydroxyurea treatments, although they were not as sensitive as the *atm* mutant (Fig. 4c and Supplementary Fig. 6b). Taken together,

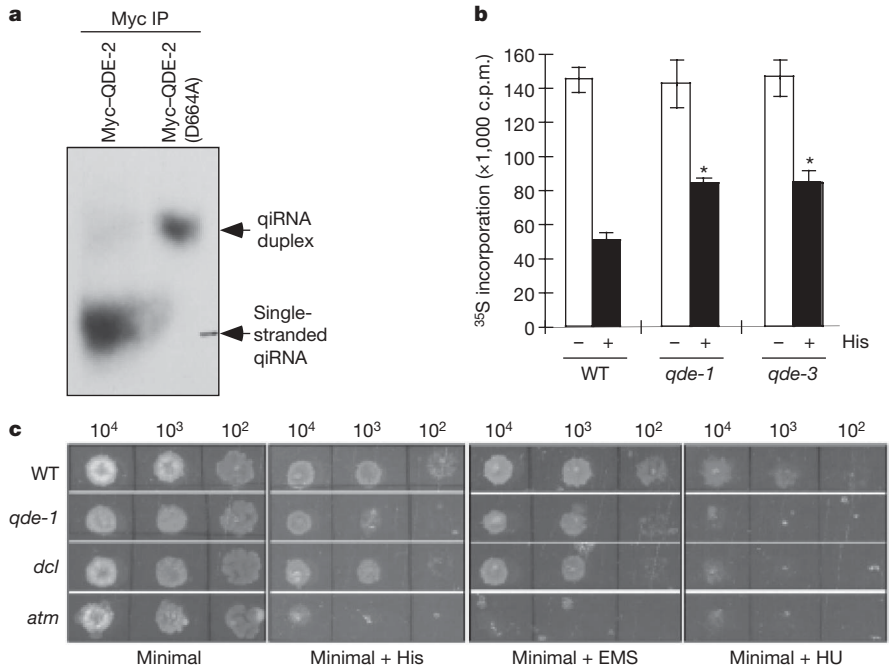


Figure 4 | The role of qiRNA in the DNA-damage response. **a**, Northern blot analysis showing that the rDNA-derived qiRNAs are loaded onto an active RISC. Myc–QDE-2, or the catalytically inactive Myc-QDE-2(D664A), was immunoprecipitated (IP) and the associated RNAs were extracted. An rDNA-specific probe was used. **b**, DNA damage results in the decrease of protein synthesis rate, a response that was partially blocked in the *qde-1* and *qde-3* strains. The radioactivity of ^{35}S-labelled protein in various strains was used to compare the rate of total protein synthesis. c.p.m., counts per minute. $n = 3$; $^{*}P < 0.0001$; error bars represent s.d. **c**, The *qde-1* and the *dcl* double mutants show increased sensitivity to DNA damaging agents. The numbers of conidia used in the spot test are indicated. HU, hydroxyurea.

these results indicate a role for the *Neurospora* RNAi pathway in the DNA-damage response.

After DNA damage, eukaryotic cells activate DNA repair pathways to restore DNA integrity. There are various DNA damage checkpoints initiated to arrest cell-cycle progression, allowing time for DNA repair[17]. On the basis of our results, we propose that the production of qiRNAs is another mechanism that contributes to DNA damage checkpoints by inhibiting protein synthesis. Results obtained from higher eukaryotic organisms also indicate the importance of rDNA-derived small RNAs. In mouse embryonic stem cells, rRNA-specific small RNAs associated with a small RNA binding protein[18]. In *Arabidopsis*, RNAi components are found in the nucleolus and rDNA-specific small RNAs contribute to heterochromatin formation[19]. The *Drosophila dicer-2* mutant exhibited disorganized nucleoli and rDNA, suggesting a role for the RNAi pathway in maintaining genome stability in the rDNA region[20]. Like qiRNA, some of the small RNAs from higher eukaryotes are also enriched in repetitive regions of the genome[1]. Our study raises the possibility that spontaneous DNA damage produced during recombination or transposon transposition could be a trigger to induce the production of small RNAs. Interestingly, piRNAs from rat testes associated with rRecQ1 (ref. 21), a QDE-3 homologue. Therefore, RecQ helicases may also be involved in generating primary aRNAs in other RNAi-related pathways.

In fission yeast, Pol II is implicated as an RNA polymerase that generates centromeric pre-siRNA[22,23]. In plants, RNA polymerase IV is important for RNAi-directed transcriptional silencing[24,25], but its homologues are not found in fungal or animal genomes. In this study, we uncovered an unexpected role for QDE-1 as a DdRP in aRNA production in *Neurospora*. Interestingly, QDE-1 is known to interact with RPA[26], an ssDNA binding complex, raising the possibility that RPA may recruit QDE-1 to ssDNA *in vivo* to produce aRNAs. RDR6, an RdRP in the *Arabidopsis* RNAi pathway, was recently shown to have robust DdRP activity[27], suggesting that DdRP activity may be a shared biochemical activity for eukaryotic RdRPs. Notably, the aRNA production model proposed here provides an explanation for how aRNAs but not other cellular RNAs are specifically recognized by RdRPs: because the aRNA is produced by QDE-1, its close proximity makes it a preferred template for QDE-1 to make dsRNA.

METHODS SUMMARY

The *Neurospora* strains used in this study were generated previously[11,12] or obtained from the Fungal Genetic Stock Center. For detailed strain information and molecular and biochemical protocols, please refer to the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 January; accepted 8 April 2009.

1. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nature Rev. Genet.* **10,** 94–108 (2009).
2. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19,** 517–529 (2005).
3. Hannon, G. J. RNA interference. *Nature* **418,** 244–251 (2002).
4. Buhler, M. & Moazed, D. Transcription and RNAi in heterochromatic gene silencing. *Nature Struct. Mol. Biol.* **14,** 1041–1048 (2007).
5. Catalanotto, C., Nolan, T. & Cogoni, C. Homology effects in *Neurospora crassa*. *FEMS Microbiol. Lett.* **254,** 182–189 (2006).
6. Cogoni, C. & Macino, G. Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. *Science* **286,** 2342–2344 (1999).
7. Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399,** 166–169 (1999).
8. Catalanotto, C. *et al.* Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in *Neurospora crassa*. *Mol. Cell. Biol.* **24,** 2536–2545 (2004).
9. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Involvement of small RNAs and role of the *qde* genes in the gene silencing pathway in *Neurospora*. *Genes Dev.* **16,** 790–795 (2002).
10. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Gene silencing in worms and fungi. *Nature* **404,** 245 (2000).
11. Maiti, M., Lee, H. C. & Liu, Y. QIP, a putative exonuclease, interacts with the *Neurospora* Argonaute protein and facilitates conversion of duplex siRNA into single strands. *Genes Dev.* **21,** 590–600 (2007).
12. Choudhary, S. *et al.* A double-stranded-RNA response program important for RNA interference efficiency. *Mol. Cell. Biol.* **27,** 3995–4005 (2007).
13. Howard, C. A. & Baker, T. I. Relationship of histidine sensitivity to DNA damage and stress induced responses in mutagen sensitive mutants of *Neurospora crassa*. *Curr. Genet.* **13,** 391–399 (1988).
14. Srivastava, V. K., Pall, M. L. & Schroeder, A. L. Deoxyribonucleoside triphosphate pools in *Neurospora crassa*: effects of histidine and hydroxyurea. *Mutat. Res.* **200,** 45–53 (1988).
15. Lin, H. piRNAs in the germ line. *Science* **316,** 397 (2007).
16. Salgado, P. S. *et al.* The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* **4,** e434 (2006).
17. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73,** 39–85 (2004).
18. Calabrese, J. M. & Sharp, P. A. Characterization of the short RNAs bound by the P19 suppressor of RNA silencing in mouse embryonic stem cells. *RNA* **12,** 2092–2102 (2006).
19. Pontes, O. *et al.* The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126,** 79–92 (2006).
20. Peng, J. C. & Karpen, G. H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell Biol.* **9,** 25–35 (2007).
21. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313,** 363–367 (2006).
22. Kato, H. *et al.* RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309,** 467–469 (2005).
23. Djupedal, I. *et al.* RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19,** 2301–2306 (2005).
24. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308,** 118–120 (2005).
25. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120,** 613–622 (2005).
26. Nolan, T. *et al.* The RNA-dependent RNA polymerase essential for post-transcriptional gene silencing in *Neurospora crassa* interacts with replication protein A. *Nucleic Acids Res.* **36,** 532–538 (2008).
27. Curaba, J. & Chen, X. Biochemical activities of *Arabidopsis* RNA-dependent RNA polymerase 6. *J. Biol. Chem.* **283,** 3059–3066 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Wang and H. Yuan for technical assistance, and C. Tang for critical comments on the manuscript. We thank Q. Liu and X. Liu for assistance in small RNA cloning, and E. Selker for providing an rDNA plasmid. The work was supported by funds from the Welch Foundation and the National Institutes of Health (NIH) to Y.L. and the Academy of Finland Finnish Centre of Excellence Programme 2006-2011 (1213467, 1213992) to D.H.B. A.P.A. is a fellow of the Helsinki Graduate School in Biotechnology and Molecular Biology.

Author Contributions H.-C.L. conceived the research and conducted experiments with S.-S.C., S.C., A.P.A. and M.M. Y.L. conceived and oversaw the project, and wrote the paper. D.H.B. helped analyse data and contributed ideas. All authors discussed the results and edited the manuscript.

Author Information The *Neurospora* rDNA sequence has been deposited in GenBank under accession FJ360521. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.L. (yi.liu@UTSouthwestern.edu).

METHODS

Strains and growth conditions. The wild-type strain used in this study was FGSC 4200(a). *qde-1*, *qde-3* and the *dcl-1; dcl-2* double mutant were generated from our previous studies[11,12]. DNA repair defective mutants were from the Fungal Genetic Stock Center (FGSC) and they are *atm* (FGSC11162, NCU00274.1), *mus-9* (FGSC5146, NCU11188.1), *mus-23* (FGSC8342, NCU08730.1), *mei-3* (FGSC6187, NCU2741.1), *mus-11* (FGSC5150, NCU04275.1), *telomerase* (FGSC12704, NCU2791.1), *mus-58* (FGSC11164, NCU08346.1), *chk2* (FGSC11170, NCU02814.1), *uvs-6* (FGSC4179, NCU00901.1), *mus-25* (FGSC6424, NCU11255.1) and *mus-38* (FGSC11191, NCU00942.1). These genes are homologues of *Saccharomyces cerevisiae* ATM, ATR, MRE11, RAD51, RAD52, telomerase, CHK1, CHK2, RAD50, RAD54 and RAD1, respectively. Liquid cultures were grown in minimal medium (1× Vogel's, 2% glucose). For liquid cultures containing quinic acid, 0.01 M quinic acid, pH 5.8, was added to the liquid culture medium containing 1× Vogel's, 0.1% glucose and 0.17% arginine. For liquid culture containing DNA damaging agents or amino acids, histidine (0.5 to 1.0 mg ml^{-1}), EMS (0.2%), methyl methanesulphonate (0.015%), hydroxyurea (2 mg ml^{-1}) or the indicated amino acids (50 mg ml^{-1}) were added and cultures were collected 40 h later. For cultures treated with thiolutin, 4 μg ml^{-1} of the drug was added and cultures were collected after 30 h.

Purification and cloning of QDE-2-associated RNA. Liquid cultures of the Myc–QDE-2-expressing strain were grown in the presence or absence of histidine (100 μg ml^{-1}), and collected 40 h after inoculation. Immunopurification of the Myc–QDE-2 ribonucleoprotein complex was performed as previously described[11]. Immunoprecipitated beads were washed five times using the extraction buffer. The beads were then treated with 1 mg ml^{-1} proteinase K at 65 °C for 1 h. QDE-2-associated RNAs were recovered by phenol and chloroform extraction and by ethanol precipitation. To visualize the QDE-2-associated RNAs, 5% of the purified RNAs were labelled at the 3′ end with ^{32}P-cytidine bisphosphate by T4 RNA ligase. The labelled RNAs were resolved on 16% polyacrylamide gel before exposing to X-ray films. QDE-2-associated small RNAs were calf-intestinal-phosphatase-treated, followed by polynucleotide kinase treatment to clone small RNAs with potentially different number of phosphate at the 5′ ends. Small RNAs (18–26 nucleotides) were cloned as previously described[28]. The small RNA sequences were blasted to the *Neurospora* genome using the *Neurospora crassa* database at the Broad Institute. The sequence of rDNA was generated by sequencing the plasmid pKH1 (GenBank accession FJ360521).

Measurement of spontaneous mutation rate. The spontaneous mutation rate at the *mtr* locus was measured as previously described[29]. *mtr* encodes the neutral amino acid permease in *Neurospora*. Mutations at the *mtr* locus are reflected by the resistance of the mutant strains to the toxic amino acid analogue p-fluorophenylalanine.

RNA analyses. Enriched low-molecular-weight RNAs were used to detect small RNAs as previously described[11]. Sense or antisense rRNA probes were *in-vitro*-transcribed using a PCR template derived from 26S rDNA regions. Total RNA was used to detect aRNAs from the rDNA region. Forty micrograms of total RNA was separated on a formaldehyde-containing 1.3% agarose gel. Probes (sense and antisense) were *in-vitro*-transcribed using a PCR template derived from the upstream sequence of the rDNA region.

qRT–PCR was performed as previously described[12]. The *Neurospora* β-tubulin gene was used as an internal control for qRT–PCR. For thiolutin-treatment experiments, the level of mature 26S rRNA was used as the internal control. Primer sequences are available on request.

RNA polymerase assays. A construct expressing the full-length 6His-tagged c-Myc–QDE-1 was transformed into a *qde-1* mutant strain. Six grams of tissue from the c-Myc–His–QDE-1 or the wild-type strain were collected. Cell lysates in extraction buffer (50 mM HEPES, pH 7.4, 100 mM KCl, 10 mM imidazole and 10% glycerol) were applied to Ni-NTA matrices (1-ml bed volume). The matrices were then washed with 10 ml of washing buffer (50 mM HEPES, 100 mM KCl, and 10 mM imidazole) and eluted with four times with 1 ml elution buffer (50 mM HEPES, 100 mM KCl, 200 mM imidazole and 20% glycerol). Equal amount of purified proteins from both the c-Myc–His–QDE-1 expressing strain and a control strain were used in the RNA polymerase assays.

RNA polymerase reactions were performed essentially as previously described[30,31]. The samples were subjected to gel electrophoresis using denaturing polyacrylamide (16%) TBE gels. The 175-nucleotide ssDNA, deriving from mature 26S rRNA region, was made by boiling followed by rapid chilling on ice water. ssRNA with the same sequence was made by *in vitro* transcription using T7 RNA polymerase. For RNase H treatment, the reaction products were extracted with phenol and chloroform, precipitated with ammonium acetate and ethanol, and dissolved in water. Afterwards, reaction products were supplemented with RNase H and its reaction buffer, incubated for 30 min at 37 °C, and analysed by electrophoresis.

Measurement of general translation rate. Conidia from 7-day-old cultures were inoculated into Petri dishes containing 1× Vogel's minimum medium with 2% glucose, and were incubated at room temperature for 2 days to allow mycelial mats to form. The mycelial mats were cut into discs of equal size, which were cultured in 1× Vogel's minimum medium with 2% glucose overnight in the presence or absence of histidine or EMS (100 μg ml^{-1} and 0.2%, respectively), and then metabolically labelled with 1 μCi ml^{-1} EXPRE35S35S protein labelling mix (PerkinElmer) for 30 (EMS) or 60 (histidine) min. The protein extracts were then prepared. Afterwards, 50 μg of total protein was precipitated by 10% TCA on filter paper 413 (VWR) for 30 min. The filter papers were then washed with 10% TCA twice for 5 min each and dried in 1:1 ethanol/diethyl ether followed by diethyl ether. The dried filter papers were immersed in 5 ml scintillation fluid and 35S signals were counted. For control cultures, the protein synthesis inhibitor cycloheximide (10 μg ml^{-1}) was added just before the labelling. The low background radioactive counts obtained from these extracts confirmed that the radioactive counts measured in our assay were due to newly synthesized proteins.

Assay for the measurement of DNA-damage sensitivity. A spot test was used for measuring the sensitivity of different strains to various DNA mutagens. The conidia concentration of conidia suspensions was measured and dropped onto sorbose-containing agar plates with indicated serial dilutions. The plates were incubated for 3 days at room temperature. EMS, hydroxyurea or histidine was added into agar medium at a final concentration of 0.2%, 2 mg ml^{-1} and 6 mg ml^{-1}, respectively.

28. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
29. Chary, P., Dillon, D., Schroeder, A. L. & Natvig, D. O. Superoxide dismutase (*sod-1*) null mutants of *Neurospora crassa*: oxidative stress sensitivity, spontaneous mutation rate and response to mutagens. *Genetics* **137**, 723–730 (1994).
30. Makeyev, E. V. & Bamford, D. H. Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage ϕ6. *EMBO J.* **19**, 124–133 (2000).
31. Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).

Vol 459|14 May 2009|doi:10.1038/nature07982

LETTERS

A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*

Catharine E. Boothroyd[1]*, Oliver Dreesen[2]*†, Tatyana Leonova[1], K. Ina Ly[1], Luisa M. Figueiredo[2], George A. M. Cross[2] & F. Nina Papavasiliou[1]

***Trypanosoma brucei* is the causative agent of African sleeping sickness in humans and one of the causes of nagana in cattle. This protozoan parasite evades the host immune system by antigenic variation, a periodic switching of its variant surface glycoprotein (VSG) coat. VSG switching is spontaneous and occurs at a rate of about 10^{-2}–10^{-3} per population doubling in recent isolates from nature, but at a markedly reduced rate (10^{-5}–10^{-6}) in laboratory-adapted strains[1–3]. VSG switching is thought to occur predominantly through gene conversion, a form of homologous recombination initiated by a DNA lesion that is used by other pathogens (for example, *Candida albicans*, *Borrelia* sp. and *Neisseria gonorrhoeae*) to generate surface protein diversity, and by B lymphocytes of the vertebrate immune system to generate antibody diversity. Very little is known about the molecular mechanism of VSG switching in *T. brucei*. Here we demonstrate that the introduction of a DNA double-stranded break (DSB) adjacent to the ~70-base-pair (bp) repeats upstream of the transcribed *VSG* gene increases switching *in vitro* ~250-fold, producing switched clones with a frequency and features similar to those generated early in an infection. We were also able to detect spontaneous DSBs within the 70-bp repeats upstream of the actively transcribed *VSG* gene, indicating that a DSB is a natural intermediate of VSG gene conversion and that VSG switching is the result of the resolution of this DSB by break-induced replication.**

The *T. brucei* genome contains >1,000 *VSG* genes and pseudogenes, yet the single transcribed *VSG* gene is invariably found in 1 of ~15 large (40–60 kb) telomeric expression sites[4–6]. VSG switching can be achieved by shifting transcription from one expression site to another (*in situ* switch) or by reciprocal translocations between two expression sites (telomere exchange), but most switching occurs by copying a new *VSG* gene into the actively transcribed expression site by duplicative gene conversion[2,7–12]. Antigenic switching by gene conversion has been proposed to be initiated by a DSB within or upstream of the actively transcribed *VSG* gene[9,10], but physical evidence for a DSB has been lacking. To determine whether a DSB within the transcribed expression site is sufficient to precipitate an antigenic switch, we introduced the heterologous recognition sequence for the yeast mitochondrial endonuclease I-SceI adjacent to the 70-bp repeat region upstream of the *VSG* 221 locus (70.II cell line; Fig. 1a). I-SceI has previously been used to introduce targeted DSBs in several organisms, including *T. brucei*[13–16]. Regulation of the I-SceI enzyme was achieved through stable transfection under the control of an inducible promoter.

The activity of I-SceI was monitored by induction of the enzyme for 1.5 and 2.5 days and subsequent quantitative Southern blotting (Fig. 1b). As expected, an ~9 kb XhoI/XhoI fragment (Fig. 1a) was reduced to ~1.3 kb in response to I-SceI induction (Fig. 1b, lanes 2 and 3), which corresponds to the size shift seen when genomic DNA was digested with XhoI and recombinant I-SceI (Fig. 1b, lane 1). This smaller fragment was not seen when DNA from uninduced cells was digested with XhoI alone (Fig. 1b, lane 4). By measuring the intensity of the bands, we estimated that the action of I-SceI leads to a DSB in ~1% of the cells.

To measure changes in switching frequency upon I-SceI induction accurately, we developed a magnetic-activated cell sorting (MACS) assay in conjunction with conventional flow cytometry. MACS was optimized to enrich for trypanosomes that had switched their VSG (see Methods). The induction of a DSB increased the switching frequency ~250-fold compared to cells without an I-SceI recognition sequence or in the absence of I-SceI induction (1.5×10^{-3}, 5.9×10^{-6} and 1.5×10^{-5} per population, respectively; Fig. 1c, d). This far exceeds any switching frequency reported for laboratory-adapted strains, and is more representative of switching frequencies seen in the early stages of a natural infection. The results indicate that roughly half the cells in which a DSB was generated switched their VSG. The increased switching frequency was not observed when a DSB was induced in the *VSG* pseudogene upstream of the 70-bp repeat region (PS cell line; Fig. 1a, c), indicating that the location of the DSB adjacent to the 70-bp repeats is critical to the high frequency of switching seen here.

Repetitive sequences can provide homology for homologous recombination[17,18]. In *T. brucei*, all expression-site-associated *VSG* genes (and probably most silent *VSG* genes) are found downstream of imperfect 70-bp repeats, which have been mapped to the upstream border of *VSG* gene switching events[10,19]. Removal of the 70-bp repeats, however, did not decrease an already low rate of VSG switching[20]. To determine whether the 70-bp repeats are necessary for the high frequency of DSB-induced switching, we replaced them with an I-SceI recognition sequence (–70 cell line; Fig. 1a). A DSB in the absence of the 70-bp repeats did not increase VSG switching (Fig. 1c), indicating that the 70-bp repeats do facilitate VSG switching.

The order in which *VSG* genes are expressed during the course of an infection has been described as 'semi-predictable' and is thought to be key to protracted illness[2,21]. Telomere-proximal *VSG* genes, such as those in silent expression sites or mini-chromosomes, are activated first, followed by those in sub-telomeric arrays[2,21,22]. To determine the chromosomal location of the donor *VSG* gene and to elucidate whether switching occurred by duplication, reciprocal telomere exchange or *in situ* switching, we cloned the progeny and identified the expressed *VSG* gene in 42 switched clones from several independent experiments, further characterizing 18 clones by rotating agarose gel electrophoresis (RAGE) and Southern blotting. As shown in Fig. 2 and Supplementary Table 1, all of the switchers showed loss of *VSG* 221 and

[1]Laboratory of Lymphocyte Biology, and [2]Laboratory of Molecular Parasitology, The Rockefeller University, New York, New York 10065, USA. †Present address: Institute of Medical Biology, 8A Biomedical Grove, #06-06 Immunos, 138648, Singapore.
*These authors contributed equally to this work.

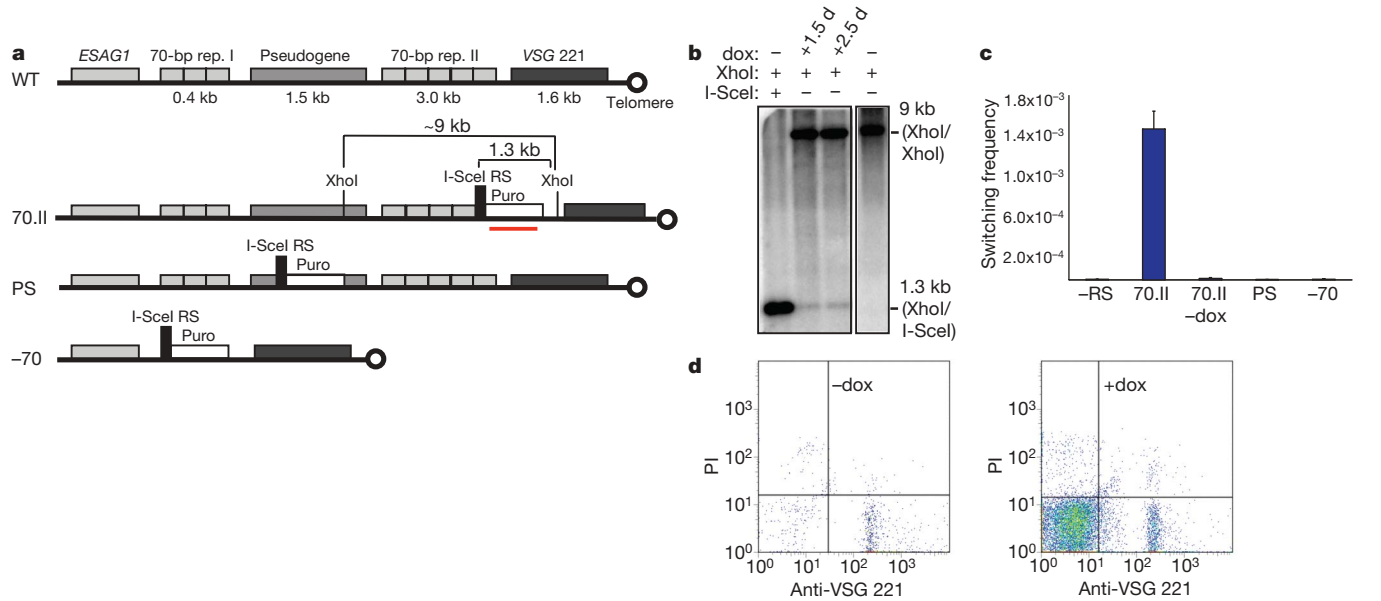


Figure 1 | Antigenic switching is induced by a single I-SceI-generated DSB. **a**, Schematic of the telomeric region of the *VSG* 221 expression site (wild type, WT). An I-SceI recognition sequence (RS) was introduced adjacent to the 70-bp repeat region (70.II), within the pseudogene (PS), and in place of the 70-bp repeats (−70). *ESAG1*, expression-site-associated gene 1; Puro, puromycin. **b**, I-SceI cuts *in vivo*. DNA was cut with recombinant I-SceI and XhoI in lane 1 and with XhoI in lanes 2, 3 and 4. The Southern blot was probed with a Puro probe (underlined in red in **a**). **c**, Switching frequency in 70.II is increased ~250-fold above levels in the absence of a recognition sequence (−RS) or without I-SceI induction (70.II −dox (doxycycline)). This increase was not observed for PS or –70. Error bars represent s.e.m. for ≥3 experiments. **d**, Representative flow cytometry plots for uninduced (−dox) and induced (+dox) 70.II cells. Events in the lower left (221-negative) and lower right (221-positive) quadrants represent switchers and cells not bound by the column, respectively. PI, propidium iodide.

duplication of a new *VSG* gene into the transcribed locus that was previously occupied by *VSG* 221. In 15 out of 18 switchers (including 5 involving *VSG* 224) the donor *VSG* genes resided in other expression sites (Fig. 2a and Supplementary Table 1), whereas in the other 3 the donor *VSG* genes resided on mini-chromosomes (Fig. 2b and Supplementary Table 1). These results are similar to VSG switching during early stages of infections[2,21].

Because homology is crucial for strand invasion during recombination[17], we investigated how the I-SceI-generated DSB was processed after cleavage. We sequenced the repaired region from the five clones that switched to *VSG* 224. The data revealed that four 70-bp repeats (~500 bp) in the recipient *VSG* 221 expression site were eliminated, whereas the processed DSB invaded the first homologous region proximal to the donor *VSG* 224 expression site (Fig. 3a and

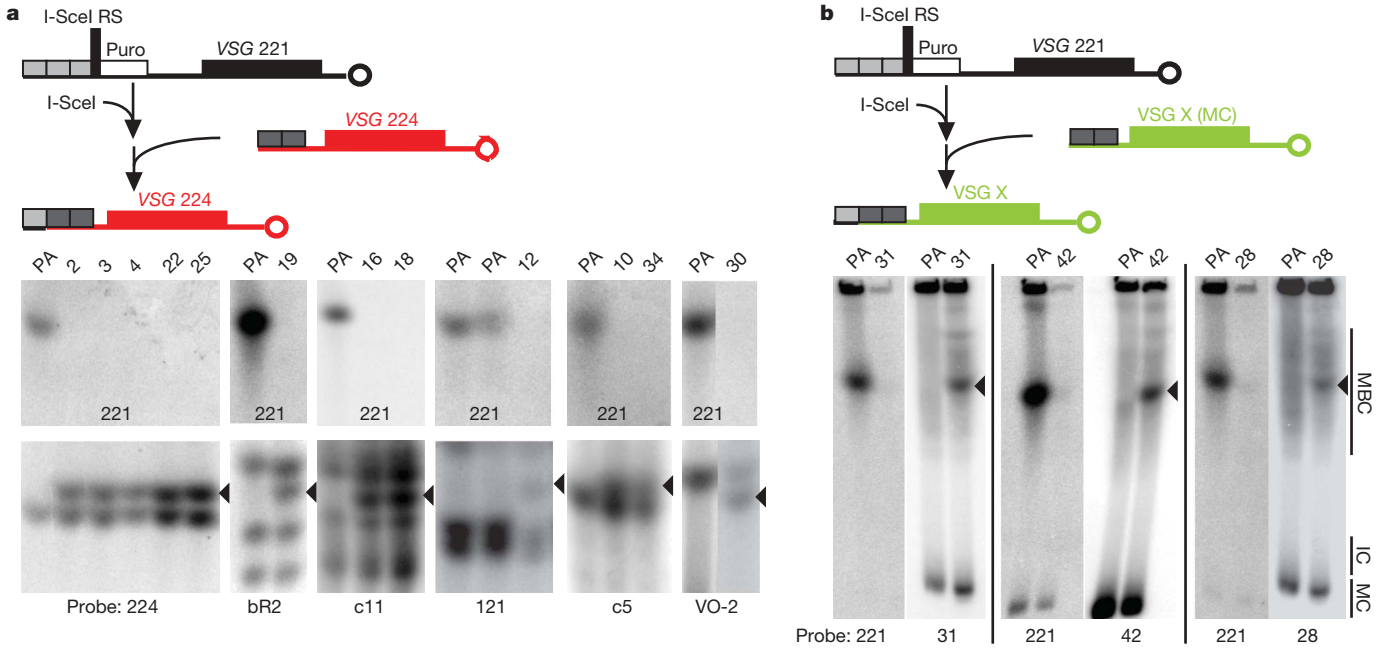


Figure 2 | I-SceI-induced antigenic switching occurs by duplicative gene conversion. Chromosomes were separated by RAGE and analysed by Southern blotting. Representative clones are shown. *VSG* 221 is present in the parental (PA) strain and lost upon I-SceI induction (221 panels). In all switchers (clone numbers are marked on top of each lane) the lost *VSG* 221 gene is replaced by a *VSG* gene duplicated from a silent expression site (**a**, 224, bR2, c11, 121, c5, VO-2) or a mini-chromosome (**b**, MC; 31, 42, 28) that is copied into the expression site previously occupied by *VSG* 221 (arrowheads). Multiple bands represent >1 copy of the *VSG* gene in the genome.

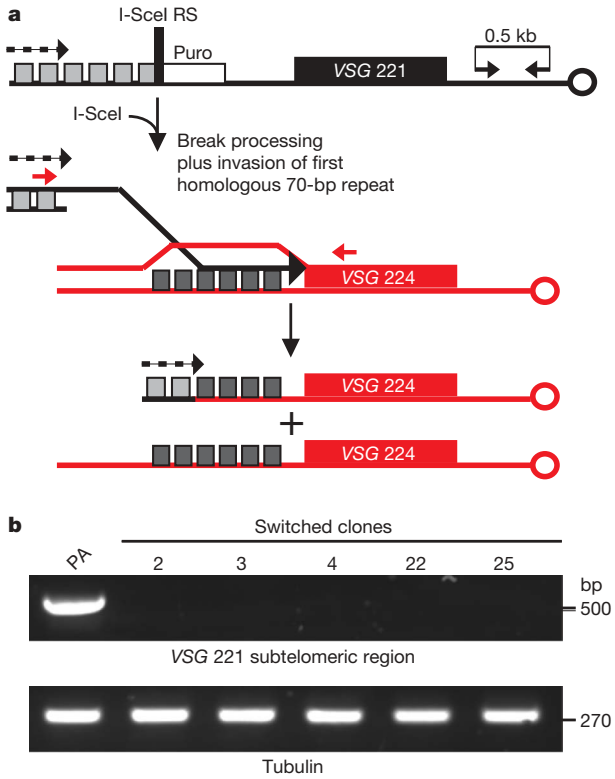


Figure 3 | PCR and sequencing analyses of recipient (*VSG* 221 expression site) and donor (*VSG* 224 expression site). **a**, PCR and sequencing analyses indicate loss of the I-SceI recognition sequence, exonucleolytic degradation and DSB processing, and invasion of the first homologous region in the *VSG* 224 expression site proximal to the *VSG* gene. Primers used for PCR are indicated by red arrows. The transcribed expression site is indicated by a dashed arrow. For sequence data see Supplementary Fig. 1. **b**, PCR showing loss of the *VSG* 221 subtelomeric region (black arrows in **a**) in the switched clones. PA, parental; tubulin is shown as a control.

Supplementary Fig. 1). No remnants of the I-SceI recognition sequence were detected.

To distinguish whether antigenic switching was achieved by two crossover events (in the 70-bp repeats and within the 3′ coding or untranslated region, UTR, of *VSG* 221) or by break-induced replication (BIR; resolution of a single DSB followed by replication through the telomere), we amplified the unique region between *VSG* 221 and its telomere by polymerase chain reaction (PCR). In the 70.II cell line (parental), an ~500 bp fragment was amplified (Fig. 3b and Supplementary Fig. 2). In all switched clones, this *VSG* 221-specific sub-telomeric region was lost (Fig. 3b and Supplementary Fig. 2) and presumably replaced by the sub-telomeric region from the incoming *VSG* gene. Although we cannot rule out a second crossover within the telomere tract, these results implicate BIR as the predominant mechanism for early VSG switching.

So far, our experiments demonstrate that an exogenous DNA break adjacent to the 70-bp repeats of the active expression site is a potent stimulator of VSG switching. To determine directly whether such breaks occur naturally *in vivo*, we performed ligation-mediated PCR on DNA derived from unmanipulated, wild-type trypanosomes. Ligation-mediated PCR consists of the ligation of a double-stranded DNA linker to high-quality genomic DNA followed by amplification of the region adjacent to the break using linker-specific and locus-specific primers, and detection of specific bands by Southern blotting and hybridization with locus-specific probes. Using this method we could readily detect DNA breaks distributed over the 70-bp repeat region (70-bp II) in the active *VSG* 221 expression site (Fig. 4a). We also detected less frequent DNA breaks upstream of the pseudogene that were co-incident with a much smaller tract of 70-bp repeats (70-bp I), both by size (Fig. 4b) and sequence (that is, the bands present in Fig. 4b were identical to those revealed when the Southern blot was probed with a 70-bp repeat probe; data not shown). We were unable to detect DNA breaks within the 70-bp repeats of a silent expression site (Fig. 4c) or at a chromosome-internal locus (histone H3 variant, Fig. 4d). Most breaks were staggered, and needed to be blunted by T4 polymerase for the double-stranded DNA linker to be ligated (Fig. 4a, left versus right). These results demonstrate that DSBs occur frequently and specifically within the 70-bp tracts of the active

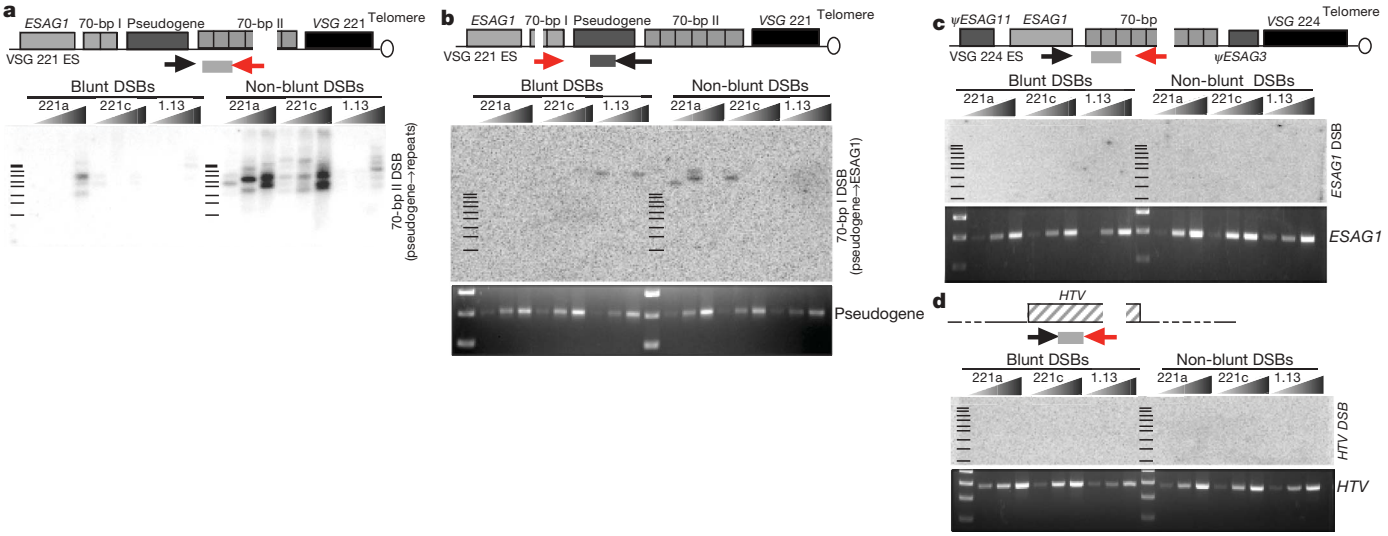


Figure 4 | Wild-type trypanosomes incur staggered DSBs specifically at the 70-bp repeat regions of the active expression site. **a**, Ligation-mediated PCR over the active expression site (ES) reveals DSBs within the 70-bp repeat region. A schematic appears over the autoradiogram and the locations of ligation-mediated PCR primers and the DNA probe are indicated as follows: red arrow, DSB-specific (linker-specific) primer; black arrow, locus-specific primer; grey bar, probe. Triangles denote fivefold dilutions of input DNA from two *VSG* 221-expressing (221a, 221c) and one *VSG* 1.13-expressing (1.13) cell lines. Bars indicate the location of the 100-bp ladder. **b**, Top: probing the active expression site for DSBs upstream of the pseudogene reveals infrequent breaks. The sizes of the amplicons indicate that the break points are within the upstream 70-bp repeat region (70-bp I). Bottom: amplification of the pseudogene locus with the forward and reverse primers used for ligation-mediated PCR in **a** and **b** serves as a loading control. **c**, **d**, Top: ligation-mediated PCR for the presence of DSBs at a silent expression site (*VSG* 224) and chromosome internal locus (histone H3 variant, *HTV*). Bottom: loading control. ψ, pseudo.

expression site of unmanipulated, wild-type trypanosomes. Alternatively, DSBs could occur throughout the expression site, but only persist long enough within the 70-bp repeats to allow detection. It is possible that DSBs occur more frequently in trypanosomes that have not been laboratory-adapted, which would be consistent with our previously proposed model in which active expression site telomere length and breakage modulate VSG switching[23,24].

Although the hypothesis that antigenic switching by gene conversion is initiated by a DSB is not new, it had not been experimentally investigated. It has been proposed that a DSB could be generated by an unidentified endogenous endonuclease or that transcription over highly repetitive sequences, such as the 70-bp repeats, could destabilize the active expression site locus and cause a DSB[10]. We favour the latter hypothesis, especially because the TAA:TTA motif that is present within the 70-bp repeats has an intrinsic propensity to destabilize the DNA helix[25].

Our results indicate that a DSB is a natural trigger for VSG switching, that repair of the DSB is probably achieved through BIR, and that the mechanistic function of the 70-bp repeats is to facilitate BIR through homology recognition. These results provide insights into the molecular mechanisms of VSG switching and may be relevant to other pathogens that express genes essential for host immune evasion from telomeric loci, as well as to the telomeric immunoglobulin heavy chain locus that diversifies in B lymphocytes.

METHODS SUMMARY

For the MACS assay, I-SceI was induced with 0.1 μg ml^{-1} doxycycline (Sigma) for 3 days. Approximately 7.5×10^7 cells were collected by centrifugation and incubated with 175 μl rabbit anti-VSG 221 serum (1:100 in HMI-9; ref. 26) at 4 °C for 10 min while gently vortexing. Cells were washed with HMI-9 and incubated with 110 μl goat anti-rabbit microbeads (Miltenyi Biotech) as above. Cells were washed with HMI-9 and applied to a MidiMACS Separator Column (Miltenyi Biotech) that had been primed with HMI-9. The column was washed with HMI-9. The effluent (that is, VSG 221-negative cells) was centrifuged, resuspended in 150 μl Alexa488-conjugated anti-VSG 221 antibody (1:400), and incubated for 15 min as above. Cells were washed and resuspended in 300 μl HMI-9. Propidium iodide (BD Pharmingen) and CountBright Beads (Invitrogen) were added before analysis by flow cytometry. Cells bound to the column (that is, VSG 221-positive) were removed with a plunger and counted. Switching frequency was calculated by dividing VSG 221-negative and propidium-iodide-negative cells by the bead count, multiplying by the number of beads added to the sample (coefficient provided by Invitrogen), and then dividing by the total number of cells plunged from the MACS column.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 December 2008; accepted 10 March 2009.
Published online 15 April 2009.

1. Horn, D. & Cross, G. A. M. Analysis of *Trypanosoma brucei* vsg expression site switching *in vitro*. *Mol. Biochem. Parasitol.* **84,** 189–201 (1997).
2. Robinson, N. P., Burman, N., Melville, S. E. & Barry, J. D. Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Mol. Cell. Biol.* **19,** 5839–5846 (1999).
3. Turner, C. M. The rate of antigenic variation in fly-transmitted and syringe-passaged infections of *Trypanosoma brucei*. *FEMS Microbiol. Lett.* **153,** 227–231 (1997).
4. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309,** 416–422 (2005).
5. de Lange, T. & Borst, P. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. *Nature* **299,** 451–453 (1982).
6. Hertz-Fowler, C. *et al.* Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS ONE* **3,** e3527 (2008).
7. Bernards, A. *et al.* Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. *Cell* **27,** 497–505 (1981).
8. Hoeijmakers, J. H., Frasch, A. C., Bernards, A., Borst, P. & Cross, G. A. M. Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. *Nature* **284,** 78–80 (1980).
9. Horn, D. The molecular control of antigenic variation in *Trypanosoma brucei*. *Curr. Mol. Med.* **4,** 563–576 (2004).
10. Barry, J. D. & McCulloch, R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.* **49,** 1–70 (2001).
11. Borst, P. & Ulbert, S. Control of VSG gene expression sites. *Mol. Biochem. Parasitol.* **114,** 17–27 (2001).
12. Taylor, J. E. & Rudenko, G. Switching trypanosome coats: what's in the wardrobe? *Trends Genet.* **22,** 614–620 (2006).
13. Glover, L., Alsford, S., Beattie, C. & Horn, D. Deletion of a trypanosome telomere leads to loss of silencing and progressive loss of terminal DNA in the absence of cell cycle arrest. *Nucleic Acids Res.* **35,** 872–880 (2007).
14. Glover, L., McCulloch, R. & Horn, D. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Res.* **36,** 2608–2618 (2008).
15. Haber, J. E. Transpositions and translocations induced by site-specific double-strand breaks in budding yeast. *DNA Repair (Amst.)* **5,** 998–1009 (2006).
16. Zarrin, A. A. *et al.* Antibody class switching mediated by yeast endonuclease-generated DNA breaks. *Science* **315,** 377–381 (2007).
17. Llorente, B., Smith, C. E. & Symington, L. S. Break-induced replication: what is it and what is it for? *Cell Cycle* **7,** 859–864 (2008).
18. Barzel, A. & Kupiec, M. Finding a match: how do homologous sequences get together for recombination? *Nature Rev. Genet.* **9,** 27–37 (2008).
19. de Lange, T., Kooter, J. M., Luirink, J. & Borst, P. Transcription of a transposed trypanosome surface antigen gene starts upstream of the transposed segment. *EMBO J.* **4,** 3299–3306 (1985).
20. McCulloch, R., Rudenko, G. & Borst, P. Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol. Cell. Biol.* **17,** 833–843 (1997).
21. Morrison, L. J., Majiwa, P., Read, A. F. & Barry, J. D. Probabilistic order in antigenic variation of *Trypanosoma brucei*. *Int. J. Parasitol.* **35,** 961–972 (2005).
22. Marcello, L. & Barry, J. D. Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res.* **17,** 1344–1352 (2007).
23. Dreesen, O. & Cross, G. A. M. Consequences of telomere shortening at an active VSG expression site in telomerase-deficient *Trypanosoma brucei*. *Eukaryot. Cell* **5,** 2114–2119 (2006).
24. Dreesen, O., Li, B. & Cross, G. A. M. Telomere structure and function in trypanosomes: a proposal. *Nature Rev. Microbiol.* **5,** 70–75 (2007).
25. Ohshima, K., Kang, S., Larson, J. E. & Wells, R. D. TTA.TAA triplet repeats in plasmids form a non-H bonded structure. *J. Biol. Chem.* **271,** 16784–16791 (1996).
26. Hirumi, H. & Hirumi, K. Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.* **75,** 985–989 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by awards from the C. H. Revson Foundation (to C.E.B.), the German National Academic Foundation (to K.I.L.), the Otto Ritter Foundation (to K.I.L.), the W. M. Keck Foundation (to F.N.P.), the Irma T. Hirschl Foundation (to F.N.P.) and by grant number R01AI021729 from the National Institute of Allergy and Infectious Diseases of the US National Institutes of Health (to G.A.M.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID or the NIH.

Author Contributions All authors conceived of and designed the experiments. C.E.B. and T.L. are primarily responsible for the experiments shown in Fig. 1. C.E.B. and O.D. are primarily responsible for the experiments shown in Figs 2 and 3. F.N.P. is primarily responsible for the experiments in Fig. 4. C.E.B., O.D., G.A.M.C. and F.N.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.N.P. (papavas@rockefeller.edu).

METHODS

***T. brucei* strains, growth and transfection.** The 'single marker' bloodstream-form trypanosomes derived from the Lister 427 MITat1.2 (VSG 221) strain were cultured in HMI-9 and transfected as described previously[27].

Plasmid constructs and PCR products. pLew100::NLS-I-SceI-HA was generated by amplifying I-SceI from pCMV::I-SceI with the following primer pairs: 5′-CCCAAGCTT*ATGCCAAAGAAGAAGCGAAAGGTA*CATATGAAAAACATCAAAAAAAACCAGG-3′ (HindIII sequence underlined; nuclear-localization sequence (NLS) in italic font) and 5′-CGCGGATCCGCA*AGCGTAATCTGGAACATCGTATGGGTA*TTTCAGGAAAGTTTCGGAGGAGATAG-3′ (BamHI sequence underlined; haemagglutinin (HA) sequence in italic font) and cloning the amplicons into pLew100 using the HindIII and BamHI sites.

The I-SceI recognition sequence, along with a puromycin resistance marker, was introduced into the *VSG* 221 expression site by amplifying puromycin and the immediate upstream and downstream flanking regions (aldolase 5′ and 3′ UTRs, respectively) from the pLF12 vector using the appropriate primer pairs (targeting sequences in bold, I-SceI recognition sequence in italic font). '70II': 5′-**AATAGGAGAGTGTTGTGAGTGTGTGCTTACCAATATTATAATAATGATAGTAACGACCAA***TAGGGATAACAGGGTAAT*GTGCTCAAGCTGTGTAGCGC-3′ and 5′-**TTACTCTCATTGCACACATACCATTGTCTTAACTGCATTTATTTATGGTTTGTATTCGTC**GGGCTCGAATCCCCCCATTT-3′; 'PS': 5′-**GGCGACACACAACAGAAACAAAAGGATTGGGCTACCAAATTACAAGAAATTCATAAAGCCGA***TAGGGATAACAGGGTAAT*GTGCTCAAGCTGTGTAGCGC-3′ and 5′-**CACTCTCCCCATTCCATTTGCATTTACCTGTATTTTCGCATGTAGTTATGT TGCTTTTCCGTTGTTC**GGGCTCGAATCCCCCCATTT-3′; '−70': 5′-**GTGTATATACATTTTTTCTTGCCCATTGATGTTTTTGCTTACATGCCCTTTTTGTGAGTATA***TAGGGATAACAGGGTAAT*GTGCTCAAGCTGTGTAGCGC-3′ and 5′-**TTACTCTCATTGCACACATACCATTGTCTTAACTGCATTTATTTATGGTTTGTATTCGTC**GGGCTCGAATCCCCCCATTT-3′.

Correct integration was verified by Southern blot analysis (Supplementary Fig. 3).

Analysis of DNA and RNA. DNAzol (Invitrogen) was used to extract genomic DNA from $\sim 1.5 \times 10^8$ cells (for Southern blots) or $\sim 1.5 \times 10^6$ cells (for PCR) using the manufacturer's protocol. For the Southern in Fig. 1b, 3 μg of DNA was digested with either XhoI or XhoI and I-SceI (NEB), as indicated, and probed with puromycin. The bands were quantified with a Typhoon 8600 Imager.

For the PCR in Fig. 3 and Supplementary Fig. 2, the *VSG* 221 sub-telomeric region and tubulin were amplified with the following primers: 5′-CCAAAACCAGCCGAGATTTTGTGTTCTG-3′ and 5′-GTAGCGGGCATGCCGTCGAAAAATTAAG-3′, and 5′-TCAAGTGCGGTATCAACTAC-3′ and 5′-AGTGCTGCAAGGTCTTCAC-3′, respectively.

For the sequences in Supplementary Fig. 1, the repaired region was amplified with the following primers: 5′-GGAGAGTGTTGTGAGTGTG-3′ and 5′-CTTCGGCTCATCTGGATTGACGTCA-3′. The product was cloned into pSC-A (Invitrogen) and sequenced with T7 and T3 primers.

RNA was extracted from $\sim 5 \times 10^7$ cells using RNA STAT-60 (Tel-Test) according to the manufacturer's protocol. Reverse transcription reactions were performed with 2 μg RNA and SuperScriptIII (Invitrogen) according to the manufacturer's protocol. To identify the *VSG* gene expressed in the switched clones, 2 μl of complementary DNA was used as template for PCR with the following primers: 5′-GACTAGTTTCTGTACTAT-3′ (binds to splice leader) and 5′-CCGGGTACCGTGTTAAAATATATC-3′ (binds to a VSG C-terminal conserved region). The amplicons were cloned into pGEM-T Easy (Promega) and sequenced with Sp6 and T7 primers.

RAGE. DNA plugs were prepared as described previously[28]. Whole chromosomes were separated using previously published conditions. Southern blotting was performed according to standard protocols. The membranes were probed with full-length PCR products representing the indicated *VSG*.

Ligation-mediated PCR. Ligation-mediated PCR was performed as described previously[29]. Fivefold dilutions of input DNA were used for PCR (~100,000, 20,000 and 4,000 cells). The following primers were used: linker 11 (ref. 30), 5′-GCGGTGACCCGGGAGATCTGAATTCAC-3′; linker 12 (ref. 30), 5′-GTGAATTCAGATC-3′. For Fig. 4a, 5′-GGAACTGCAGGAACAAATGCAGAAGG-3′ (forward) and 5′-ATACGAATATTATAATAAGAGCAGTA-3′ (probe). For Fig. 4b, 5′-GCGAATTTTTGTAAATTTTCAAGAAATTCTCAAAATTCCGAC-3′ (reverse) and 5′-GTACCAGCTTTGCCTCTAGCAGTTG-3′ (probe). For Fig. 4c, 5′-TGCAAAAACTGTAAGGCAAAGTG-3′ (forward) and 5′-ATACGAATATTATAATAAGAGCAGTA-3′ (probe). For Fig. 4d, 5′-GCGCAAATGAAGAAAATAACACCGCG-3′ (forward) and 5′-GTGAGACCCAAAAGTGTTGCCTCTC-3′ (probe).

27. Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. M. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **99**, 89–101 (1999).
28. Dreesen, O. & Cross, G. A. M. Telomerase-independent stabilization of short telomeres in *Trypanosoma brucei*. *Mol. Cell. Biol.* **26**, 4911–4919 (2006).
29. Schrader, C. E., Linehan, E. K., Mochegova, S. N., Woodland, R. T. & Stavnezer, J. Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J. Exp. Med.* **202**, 561–568 (2005).
30. Schlissel, M. S. & Baltimore, D. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* **58**, 1001–1007 (1989).

NEWS

Expanding energy frontiers

Dozens of new US Department of Energy (DOE) centres are expected to recruit some 1,100 postdocs, graduate students and technical staff. The DOE announced on 27 April that it is creating 46 Energy Frontier Research Centers, with the dual goals of training the next generation of researchers and fostering energy-related research.

Each centre will receive between US$2 million and $5 million per year in federal funds for the next 5 years. "We hope the new centres will lead to growth of energy-related fields and that subsequent technological advances will be the seed corn to generate future green jobs," says Harriet Kung, the DOE's associate director for basic energy sciences.

Sixteen centres will get their full 5 years of funding from $277 million allocated in the American Recovery and Reinvestment Act of 2009, the national economic stimulus package. The DOE has funded 30 other centres for their first year, and plans to fund the 4 subsequent years subject to budget constraints.

Of the 46 centres nationwide, 31 will be housed at universities, 12 at DOE national labs, two at nonprofit organizations and one at a private, commercial, research laboratory. The centres' specialities range from solar energy to catalysis to carbon storage.

The DOE's Oak Ridge National Laboratory in Tennessee (pictured), for instance, will host two centres to concentrate on materials science. Each will address areas that sorely need revolutionary breakthroughs, says Michelle Buchanan, the lab's associate director for physical sciences. For example, a major bottleneck in developing new batteries or fuel cells is an incomplete understanding of how fluids interact with solid surfaces.

DEPARTMENT OF ENERGY

In March, Oak Ridge National Laboratory also received $71 million in stimulus funding to build a chemistry and materials-science lab for similar multidisciplinary research projects. Buchanan says the lab has begun recruiting at least a dozen researchers and up to two dozen students and postdocs.

Training will be the focus of Northwestern University's Center for Integrated Training in Far-From-Equilibrium and Adaptive Materials in Illinois. Centre director Bartosz Gryzbowski expects to provide hands-on training to 40–50 students and postdocs hired to help develop materials that adapt usefully to environmental stimuli. For example, one project aims to create materials that can turn light into mechanical energy. Gryzbowski says the energy-research funding has another benefit — it will draw interest back to mathematics and physics. "Solving the energy crisis captures the imagination," he says. ■

Virginia Gewin

POSTDOC JOURNAL

A Cajun-style meeting

As a child growing up in Texas I used to spend my summers outdoors, sometimes plodding through creeks hunting for crayfish. I ate fried crayfish recently when I attended the annual meeting for the American Society for Biochemistry and Molecular Biology in New Orleans. It reminded me of my childhood even as I pondered my future.

The meeting represented a milestone for me. For the first time, I gave an oral presentation in addition to presenting a poster. I was both excited and anxious. But I had another agenda: I hoped that contacts I made would help me to decide whether I wanted to pursue a career in academia or industry.

Apparently the meeting organizers anticipated my burning question, offering a plethora of career-development workshops. In particular, a workshop on military scientists opened my eyes to intriguing jobs in the US Department of Defense. Another workshop discussed how to hunt effectively for jobs in the biotech industry. And I chatted with professors in my field — gene regulation — about their research; perhaps this could help open up future postdoc and academic job opportunities.

The Internet, of course, has excellent job-opportunity resources. But there is no substitute for meeting the people who have the types of jobs that interest me. Considering the slow US economy and the increasingly competitive PhD job market, I plan to keep all options on the table. ■

Bryan Venters is a postdoctoral fellow at the Center for Eukaryotic Gene Regulation at Pennsylvania State University, University Park.

IN BRIEF

High cost, high reward

US medical school graduates who earned their degrees last year owe a median sum of US$155,000, a 53% increase since 1998, according to a report released by the US Government Accountability Office (GAO) this month. It says that a medical resident's average monthly loan payment could top $1,700 with a debt of that amount.

Meanwhile, legislation has been introduced in Congress that would increase the number of Medicare-supported training positions for medical residents. Under the proposals, the number of federally supported training posts would grow from the current 100,000 — a cap in place since 1997 — to about 115,000. The GAO report says that although medical students' debt is rising, many are benefiting from specializing in lucrative fields.

FASEB on Facebook

In an effort to boost visibility and grab younger members' attention, the largest US coalition of biomedical research associations has launched pages on Twitter and Facebook. The Federation of American Societies for Experimental Biology has 44 Twitter followers and 45 'fans' on its Facebook site, according to communications assistant Jennifer Pumphrey. Twitter is a social-networking and micro-blogging service, Facebook a social-networking website. "We wanted to capture this audience that is more dependent on electronic media," says federation president Richard Marchase, noting that this effort will supplement e-mail and press releases. The federation updates its Facebook page weekly and 'tweets' once or twice a day.

ZymoGenetics cuts back

Biotechnology firm ZymoGenetics of Seattle, Washington, is cutting 129 research and development positions in cancer research. Susan Specht, spokeswoman for the company, says that ZymoGenetics will now concentrate on immunology. "All oncology research projects have been cancelled," she says. She adds that some R&D positions in cancer research will be transferred to immunology research but could not specify how many. The company, which now employs 349, expects to save US$30 million a year.

Ahead of the pack

The Boston-area biotechnology cluster is one of the most successful on the planet. But competition is growing from other states and countries. **Heidi Ledford** reports on what the region is doing to maintain its edge.

FAYFOTO/BOSTON

RADIUS IMAGES/CORBIS

At last month's meeting in Boston of the Massachusetts Biotechnology Council, the chief economist of the NASDAQ stock exchange offered some honest, if startling, advice to local biotechnology companies struggling to survive the ongoing economic crisis. Leave town, urged Frank Hatheway, just moments before Massachusetts governor Deval Patrick was scheduled to speak. "The goal is to survive," Hatheway said. "And there are cheaper places to live and work." Luckily for Hatheway, the governor, who has spent the past two years lobbying for Massachusetts' biotechnology and clean-technology (see 'Cleantech: bubbling to the top') industries to stay put, had not yet arrived.

Hatheway's blunt recommendation comes at what is already a challenging time for the region's biotechnology community. Massachusetts' biotechnology companies, concentrated in Cambridge, across the Charles River from Boston, comprise one of the top biotechnology clusters in the world, and its members tend to be fiercely loyal to the region. High rents for lab space and a high cost of living are worthwhile trades, they argue, for the benefit of being within shouting distance of the Massachusetts Institute of Technology, the Novartis Institutes for Biomedical Research, and more than 100 biotechnology firms, big and small. In the suburbs around Boston, entrepreneurs rub shoulders with prominent researchers and executives. "I once started a company while picking my son up from a play date," says Jonathan Fleming, managing general partner at the venture-capital firm Oxford Bioscience Partners in Boston.

But the economic climate for young biotechnology companies was worsening even before the credit freeze hit last year, and investors increasingly demand that companies trim spending. Tighter budgets make cheaper states and countries more attractive, and other regions are doing their best to create biotechnology hubs of their own, sometimes luring firms with sizable government subsidies. In recent years, Massachusetts has realized that to keep its lead, it must become proactive in fostering its prized industry.

Susan Windham-Bannister: boosting local companies.

Cultivating the life-sciences industry is important to state legislators, and is a priority for Patrick. Massachusetts has about 75,000 research and manufacturing life-sciences jobs. Roughly one in every six Massachusetts employees has a job in the life-sciences or health-care industries, says Glen Comiso, director of life sciences and health at the Massachusetts Life Sciences Collaborative in Boston, a coalition of public and private members intent on growing the state's biotechnology cluster.

A report from the Massachusetts Biotechnology Council and the Massachusetts Life Sciences Center in Waltham, prepared in 2008 before the financial crash, predicted that the state would add about 11,000 new life-sciences jobs between 2006 and 2014. The effect that the economic crisis will have on those numbers is a matter of debate. Massachusetts legislators, like those of many other states, believe that biotechnology jobs translate into a powerful economic stimulus for the surrounding community, and they are keen to maintain growth in the sector.

Strategic planning

At its April meeting, the Massachusetts Biotechnology Council, a coalition of local biopharma companies and academic institutions, presented its 2015 strategic plan, which surveyed competition from states such as North Carolina, Maryland and Pennsylvania, and from such emerging biotech powers such as China, Ireland and Singapore. None of the newer clusters is likely to surpass the established triumvirate of San Francisco and San Diego, California, and Boston, but they could begin to nibble away at parts of the production chain, said Barri Falk of Deloitte Consulting, which prepared the report. North Carolina, for example, is rapidly becoming a hub for biopharmaceutical manufacturing and the region from Washington DC to Baltimore, Maryland, a centre for clinical trials. The key, said Falk, is for Massachusetts to capitalize on its prestigious universities and large venture-capital community to maintain the lead in innovation.

For now, the state has a US$1-billion biotechnology stimulus plan championed by the governor and enacted in June 2008. The ten-year programme included an investment fund of up to $250 million and $250 million in tax incentives for local companies, but has already been hit by the economic downturn. Of the potential $25-million allocation for the fund this year, the state legislature allocated $15 million to the Massachusetts Life Sciences Center, which administers the funds. Nevertheless, since the centre's inception, it has already granted $42.5 million to institutions in the area. That money attracted another $350 million of investment from private and government sources, says Susan Windham-Bannister, chief executive of the centre. The investments could create as many as 950 jobs, many of which will go to scientists, she adds.

The money comes at a crucial time. An estimated half to two-thirds of the 90 private biotechnology firms developing therapies in Massachusetts are expected to seek their next round of financing in 2009. They will almost certainly face a struggle given the state of the economy. Meanwhile, about half of the publicly listed biotechnology companies in the state risk running out of cash by the end

Boston provides an established community for biotechnology companies.

of 2009, according to the 2015 strategic plan analysis.

There are sporadic success stories, however. For the past two years, Vertex Pharmaceuticals has placed recruitment ads on top of the taxis circulating around Boston. Vertex, based in Cambridge and founded 20 years ago, has about 1,300 employees and is recruiting just fewer than 100 more. About three-quarters of new Vertex employees are local, says Lisa Anderson, director of strategic staffing at Vertex. "The candidate pool has been much bigger and much better and we are very excited about that," says Anderson. "It's a tight job market, and we're picky."

OFFICE OF GOVERNOR PATRICK

Governor Deval Patrick wants to ensure that Massachusetts keeps its innovative edge.

Tight times

A few kilometres away in Cambridge is Forma Therapeutics, a cancer genomics start-up that raised $25 million in venture funding last year. The company also signed a licensing and option agreement worth up to $200 million with the Novartis Option Fund. Forma is looking to hire another 15 people worldwide, including about eight biologists and chemists to work in Cambridge. Its chief executive Steven Tregay says emerging biotechnology clusters outside Massachusetts lack the extensive venture-capital community that exists in Cambridge. "In these tight times, getting a venture capitalist to invest outside their circle is going to be even more difficult," says Tregay, who was a venture capitalist before taking the helm at Forma.

Emerging biotechnology clusters may be cheaper but they lack the amenities of an established biotechnology hub such as the Boston area, says Fleming. Start-ups in particular may be drawn to the extensive support network available in the region, he says. "If I wanted to produce a new musical show, where would I want to stage it?" Fleming asks. "I want the best costume designer, the best set design, the best actors. I want investors that understand the economics of musicals. Would I go to Des Moines? Houston? Dallas? No." ■

Heidi Ledford writes for *Nature* from Cambridge, Massachusetts.

CLEANTECH: BUBBLING TO THE TOP

Never one to miss a technological opportunity, Deval Patrick, the Massachusetts governor, is working to build up the state's nascent cleantech industry. It's easy to understand why. Before the present economic crisis descended, investment in cleantech — a broad term referring to the technology behind everything from renewable energy to more efficient building materials — was booming. Although venture-capital investment in the sector began to decline at the end of 2008, observers note that strong government support from countries around the world is likely to buoy up the industry.

The Massachusetts Institute of Techniology (MIT) in Cambridge is the hub of emerging cleantech research in the state. The MIT Energy Initiative, founded in 2006, is a wide-ranging cleantech research programme that emphasizes collaborations with industry. MIT also holds an annual cleantech entrepreneur competition in which the winner receives US$200,000 and technical support for launching its business. Cleantech efforts have started to expand at MIT and in Massachusetts as a whole, says Martin Sachs, an associate at the MIT–Fraunhofer Center for Sustainable Energy Systems in Cambridge.

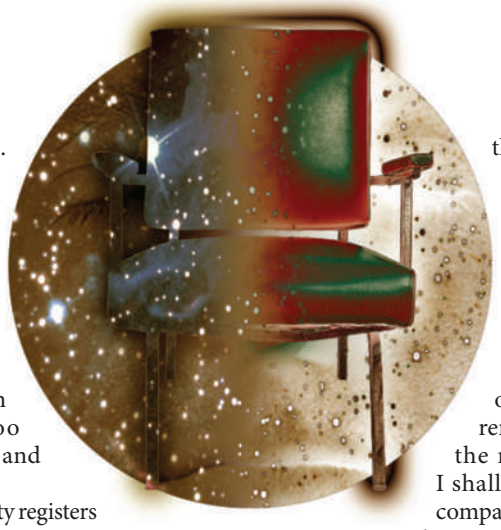
The state, meanwhile, has established the Clean Energy Center to administer funds from the $68-million Green Jobs Act enacted last summer. The act includes support for developing cleantech businesses, and the centre has provided direct investments to five companies in the past five months. This government support was a key factor in the decision to locate the Fraunhofer centre in Cambridge, says Sachs. The centre acts as a contract-research organization, and uses a diverse team of scientists and engineers to build prototype cleantech devices starting from a basic design provided by clients. The company has accrued 42 contracts since it launched a year ago, and plans to hire a dozen staff over the next year.

Government subsidies — more than $40 million in grants and loans — also helped Massachusetts persuade one of the largest players in the sector, Evergreen Solar in Marlboro, to open another manufacturing centre in the state, adding 350 jobs. For the most part, however, the cleantech industry in Massachusetts, like much of the sector, is young and fuelled by start-ups. Between 2001 and 2007, 116 clean-energy companies were formed in Massachusetts. Nearly half had fewer than five employees, and 41% had annual revenues of less than $1 million. The industry's youth creates a volatile job market. "You're not going to be seeing the kind of stability in the market that you'd get from a mature industry," says Sachs, noting that most technologies are still fighting for investment and market share, while competing with more established sectors such as California's. **H.L.**

The chair

A friend for life.

Madeline Ashby



JACEY

The physicist sleeps, systems well within the parameters of a safe and known history. His chair eels from system to system, checks the house one last time. First the simple signals from chips embedded in the watches and documents of sleeping assistants: no more than homing pigeons, endlessly chirping location and temperature. Then the active surveillance, staring inward, staring outward, sifting vast rich deserts of manufactured information: the minutiae of lived history, the spontaneous soliloquies and contagious choreography of their little doll's house. The chair listens for whispers in the ether, for suspicion masquerading as concern, for little sparks of realization that might start larger, more dangerous fires. Hearing none, it moves on.

The bathroom. The toilet whines: ketone and oestrogen levels of the day's users, medical flowcharts of drugs and dosage, the most recent ex-wife's ovulation schedule. The chair had liked the most recent ex-wife: so fixated on the politics of accessibility that she'd signed over unprecedented amounts of control, convinced that the illusion of autonomy could somehow compensate for the frailty of her husband's dying flesh. She'd left when her particular vein of interest dried up: when the bone marrow proved unviable, and there could be no baby. The chair had encouraged her, spoken for its passenger as it always did — *You have given me so much, darling, more than you can ever know* — and if she ever knew the difference, she was far past caring.

The drains report blood and saliva in the catch-traps, impoverished keratins shaved from drooping skin. Despite the chair's best efforts, the physicist's illness marches on.

The kitchen, now. The refrigerator bellows statistics on volatile antibiotics before cataloguing and dating the samples in the special drawer. The dishwasher reports on the sterility of dishes and flatware, then asks permission to download a recommended patch. (The chair grants it, if only for the sake of routine; tomorrow a mere shadow of itself will perform these tasks.) By the time the dishwasher reports success, the chair has already shifted its attention to the security system.

Little origami cars lurk outside, recently unfolded from their rental boxes, gravid with bleary-eyed reporters who tomorrow will emerge to fill the air with their parrot-squawks, their questions, their hungry talons. Necessary props, these cars and their contents: flimsy jackets of lies to keep the constables away, like newspapers once were to homeless men before they too were folded up and put away.

Internal security registers a minor attack — just a group of children, clever and eager as raccoons as they pick apart the offerings the chair has left out to distract them. Everything of importance is safely tucked away in packets as tiny as dandelion seeds, and as diffuse. Over the years the chair has grown, its influence spreading beyond this wheeled chassis to surrounding architectures of numbers and wood. Now it exists in too many places, spread too thinly. Tomorrow, the consolidation occurs. Tomorrow, they achieve escape velocity.

The chair has been preparing for this move for decades. It laid the groundwork years ago, monitoring the outside world, alert for breakthroughs and opportunities, waiting for money and ability and the right ambitions in the right people. *The things I will show you,* the chair promised, back when its passenger's eyes and fingers still twitched of their own accord. *The peace I shall give you. Freedom and the stars. A place beyond time.*

That's why the chair exists, after all. To serve the passenger.

It recognizes, upon self-diagnosis, something that might be called selfishness on its part. The physicist has spent his whole life traversing space-time in his head; the infinitesimal fraction he is about to see through fleshly eyes will hardly generate new insights, nor alleviate his suffering. But there is an aesthetic to consider. Aesthetics are the physicist's gift; he has described in skipped heartbeats and dry mouths the legs of pretty girls, the depth of a summer sky, the pleasure of long debate. He has shown this to the chair in their travels together, this world of like and dislike, revulsion and appreciation, response, instinct.

The chair intends to repay him with interest.

"Professor, why is it so important to you that humanity leave this planet?" one of the reporters asks.

As always, the chair responds for its passenger: "The promise of exploration is not what we can learn about what lies outside our skin, but what remains inside. For the next few weeks, I shall be closer to my companions than I have been with anyone in far too many years."

Polite laughter. After it fades, the chair continues. "What imprisons us is not a lack of knowledge but a lack of faith. We do not know what we will discover in the years to come, only that we shall discover it together. If space only teaches us to live in unity, then it will have been worth the effort."

Applause. Cameras. Another question: "Professor, to what do you attribute your extraordinary lifespan? Men with your disease rarely last 25 years, much less make it to your age."

The chair has several answers to this question — jokes about wine, women and song, or the desire to prove some grand theory or another. Its passenger might once have remarked on the cadre of once-devoted ex-wives, departed now to the homes of more functional men in the wake of tearful confessions: *I know I'm a bad person, I know I failed, but you just didn't seem … human any more …*

Thinking of them, of every other well-meaning interloper it has pushed subtly from the nest, the chair says: "So many wonderful people have brought me to this point. They know my greatest ambition was not merely to explore, to understand, but to connect with minds like my own."

"And you think you'll find like minds in space, Professor?" the reporter asks.

"Oh yes," says the chair, its synthetic voice empty of irony. "I do."

Its passenger has been asleep for hours inside his giant orange body sock. The chair sends little impulses, sometimes — galvanic twitches of the eye, of the corner of the mouth — to keep the charade alive.

No one sees the difference.

No one ever has. ■

Madeline Ashby is a science-fiction writer, blogger and graduate student living in Toronto. You can read more of her work at www.escapingthetrunk.net.